

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:37:18 ; Search time 3312.9 Seconds  
(without alignments)  
132.650 Million cell updates/sec

Title: US-09-235-416-3  
Perfect score: 21  
Sequence: 1 atgtcggcggtggaataatc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	18.4	87.6	143981	9	HSJ1050K3	AL121776 Human DNA
c	17.8	84.8	204067	2	AL662786	AL662786 Mus muscu
c	16.8	80.0	737	2	HS336359	AJ336359 Homo sapi
c	16.8	80.0	750	9	HS336191	AJ336191 Homo sapi
c	16.8	80.0	792	9	HS334149	AJ334149 Homo sapi
c	16.8	80.0	88941	2	AC106623	AC106623 Rattus no
c	16.8	80.0	100912	2	AC021089	AC021089 Homo sapi
c	16.8	80.0	104188	2	AF004257	AF004257 Oryza sat
c	16.8	80.0	107025	2	AL139235_3	Continuation (4 of
c	16.8	80.0	172300	2	AC106570	AC106570 Rattus no
c	16.8	80.0	194322	3	AC024791	AC024791 Caenorhab
c	16.8	80.0	319857	2	AC006782	AC006782 Caenorhab
c	16.8	80.0	57013	2	AC015769	AC015769 Homo sapi
c	16.8	80.0	63528	2	AC091196	AC091196 Homo sapi
c	16.8	80.0	100511	2	AC010774	AC010774 Homo sapi
c	16.8	80.0	156929	2	AC073751	AC073751 Mus muscu
c	16.8	80.0	196057	2	AC073725	AC073725 Mus muscu
c	16.2	77.1	3505	8	AB045127	AB045127 Schizosac
c	16.2	77.1	27777	3	AC002447	AC002447 Drosophill
c	16.2	77.1	42844	3	CEC35D6	282261 Caenorhabd
c	16.2	77.1	44751	8	SPCC16C4	AL031535 S.pombe c
c	16.2	77.1	56780	2	AC107297	AC107297 Homo sapi
c	16.2	77.1	58950	2	AC102353	AC102353 Mus muscu
c	16.2	77.1	91202	2	AC007609	AC007609 Homo sapi
c	16.2	77.1	110000	2	LMFLCHR18_04	Continuation (5 of
c	16.2	77.1	119558	2	AC097782	AC097782 Rattus no
c	16.2	77.1	138877	2	AF004398	AF004398 Oryza sat
c	16.2	77.1	140688	2	AP003801	AP003801 Oryza sat
c	16.2	77.1	148366	2	AC108717	AC108717 Homo sapi
c	16.2	77.1	151345	2	AP003332	AP003332 Oryza sat
c	16.2	77.1	152147	9	AC091842	AC091842 Homo sapi
c	16.2	77.1	157774	2	AC013353	AC013353 Trypanoso
c	16.2	77.1	163954	2	AC017155	AC017155 Drosophill
c	16.2	77.1	164558	2	AC040959	AC040959 Homo sapi
c	16.2	77.1	165507	2	AC023823	AC023823 Homo sapi
c	16.2	77.1	166941	2	AC093994	AC093994 Rattus no
c	16.2	77.1	175223	9	AC004617	AC004617 Homo sapi
c	16.2	77.1	177863	2	AL359388	AL359388 Homo sapi
c	16.2	77.1	178454	2	AC055776	AC055776 Mus muscu
c	16.2	77.1	180401	9	AC007489	AC007489 Homo sapi
c	16.2	77.1	185063	2	AC104036	AC104036 Homo sapi
c	16.2	77.1	186432	3	AC099034	AC099034 Drosophill
c	16.2	77.1	190803	9	AL355490	AL355490 Human DNA
c	16.2	77.1	302422	3	AE003455	AE003455 Drosophill
c	16.2	77.1	303040	1	AP000991	AP000991 Thermopia

ALIGNMENTS

RESULT 1

HSJ1050K3

LOCUS

DEFINITION

143981 bp DNA linear PRI 04-APR-2001  
Human DNA sequence from clone RP5-1050K3 on chromosome 20. Contains part of the EYA2 gene for eyes absent (Drosophila) homolog 2, a Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) pseudogene, a RPL27A (60S ribosomal protein L27A) pseudogene, ESTs, STSs and GSSs, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (03-APR-2001)

CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jun 3, 2000 this sequence version replaced gi:7799934.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr20>  
This sequence is the entire insert of clone RP5-1050K3. The true right end of clone RP11-323C15 is at 58476 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-1050K3 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>  
VECTOR: pCVPAC2.

## FEATURES

source	Location/Qualifiers
	1. .143981
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/clone="RP5-1050K3"
	/clone_lib="RPCI-5"
	134. .175
repeat_region	/note="L2 repeat: matches 2709. .2750 of consensus"
repeat_region	485. .664
repeat_region	/note="MIR repeat: matches 46. .260 of consensus"
repeat_region	774. .969
repeat_region	/note="MIR repeat: matches 70. .262 of consensus"
repeat_region	1091. .1330
repeat_region	/note="MER46A repeat: matches 2. .235 of consensus"
repeat_region	1387. .1679
repeat_region	/note="AluX repeat: matches 1. .294 of consensus"
repeat_region	2480. .2676
misc_feature	/note="MIR repeat: matches 41. .262 of consensus"
	complement(2633. .2991)
	/note="match: GSS: Em:AQ017978"
repeat_region	2847. .2874
	/note="14 copies 2 mer tt 89% conserved"
misc_feature	2926. .3557
	/note="match: GSS: Em:AQ423175"
misc_feature	2930. .3460
	/note="match: GSS: Em:AQ423187"
repeat_region	3022. .3103
repeat_region	/note="L2 repeat: matches 2669. .2750 of consensus"
repeat_region	3109. .3326
	/note="MIR repeat: matches 20. .246 of consensus"
repeat_region	3344. .3543
	/note="LIMC4 repeat: matches 7750. .7977 of consensus"
repeat_region	3544. .4059
repeat_region	/note="MUT1F repeat: matches 1. .541 of consensus"
repeat_region	4060. .4192
repeat_region	5002. .5192
	/note="MIR repeat: matches 7626. .7750 of consensus"
misc_feature	complement(5198. .5634)
repeat_region	/note="match: GSS: Em:AQ787200"
	5411. .5684
repeat_region	/note="MIR repeat: matches 1. .262 of consensus"
misc_feature	5646. .6090
	/note="match: GSS: Em:AQ609792"
repeat_region	6373. .6495
	/note="AluSp repeat: matches 140. .298 of consensus"
	/note="MIR repeat: matches 53. .192 of consensus"
	6517. .6723
	/note="MIR repeat: matches 4. .215 of consensus"
repeat_region	6768. .6950
	/note="MIR repeat: matches 63. .245 of consensus"
repeat_region	6951. .7253
	/note="AluJo repeat: matches 1. .305 of consensus"
repeat_region	7254. .7265
	/note="MIR repeat: matches 245. .256 of consensus"
repeat_region	7395. .7610
	/note="AluJo repeat: matches 95. .310 of consensus"
repeat_region	7809. .7952
	/note="MIR repeat: matches 28. .184 of consensus"
repeat_region	7953. .8262
	/note="AluX repeat: matches 1. .312 of consensus"
repeat_region	8263. .8288
	/note="MIR repeat: matches 184. .207 of consensus"
repeat_region	8289. .8450
	/note="MER20 repeat: matches 1. .164 of consensus"
repeat_region	8596. .8676
	/note="MIR repeat: matches 109. .199 of consensus"
repeat_region	9736. .9877
	/note="MIR repeat: matches 109. .255 of consensus"
repeat_region	10029. .10325
	/note="AluJB repeat: matches 11. .295 of consensus"
repeat_region	10370. .10506
	/note="MIR repeat: matches 103. .241 of consensus"
repeat_region	11776. .11966
	/note="MER58A repeat: matches 6. .224 of consensus"
repeat_region	12110. .12326
	/note="MER58A repeat: matches 4. .223 of consensus"
repeat_region	12716. .12835
	/note="MLT1I repeat: matches 286. .407 of consensus"
repeat_region	13032. .13336
	/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region	13374. .13594
	/note="MER58A repeat: matches 4. .224 of consensus"
repeat_region	13690. .13795
	/note="L2 repeat: matches 2583. .2705 of consensus"
misc_feature	14045. .14421
	/note="match: GSS: Em:AQ031759"
repeat_region	14143. .14337
	/note="MIR repeat: matches 1. .204 of consensus"
misc_feature	complement(14410. .14921)
	/note="match: GSS: Em:AQ055749"
repeat_region	14664. .14692
	/note="L2 repeat: matches 164. .192 of consensus"
repeat_region	14853. .14895
	/note="L2 repeat: matches 2707. .2749 of consensus"
repeat_region	14981. .15113
	/note="MIR repeat: matches 4. .144 of consensus"
repeat_region	15094. .15249
	/note="MIR repeat: matches 9. .177 of consensus"
repeat_region	15227. .15297
	/note="L2 repeat: matches 2621. .2689 of consensus"
misc_feature	complement(16294. .16547)
	/note="match: GSS: Em:B14853"
misc_feature	17097. .17504
	/note="match: GSS: Em:AQ128898"
repeat_region	17149. .17400
	/note="L2 repeat: matches 2468. .2740 of consensus"
repeat_region	17433. .17605
	/note="L2 repeat: matches 2192. .2374 of consensus"
repeat_region	17676. .17729
	/note="L2 repeat: matches 2695. .2750 of consensus"
repeat_region	18336. .18486
	/note="MIR repeat: matches 98. .261 of consensus"
repeat_region	18770. .18941
	/note="MER20 repeat: matches 3. .184 of consensus"
repeat_region	19066. .19554
	/note="L2 repeat: matches 2157. .2709 of consensus"
repeat_region	20336. .20489
	/note="AluSp repeat: matches 140. .298 of consensus"

repeat\_region 21031..21255  
/note="MER58A repeat: matches 1..224 of consensus"  
repeat\_region 21526..21609  
/note="MLT11 repeat: matches 322..410 of consensus"  
repeat\_region 21723..21925  
/note="MLT1J repeat: matches 19..261 of consensus"  
repeat\_region 21942..22258  
/note="THE1B repeat: matches 1..364 of consensus"  
repeat\_region 22267..22337  
/note="MIR repeat: matches 180..250 of consensus"  
repeat\_region 22535..23043  
/note="MLT1D repeat: matches 7..497 of consensus"  
misc\_feature 23092..23310  
/note="match: GSS: Em:AQ306389"  
repeat\_region 23762..24321  
/note="L1PA13 repeat: matches 5573..6142 of consensus"  
misc\_feature complement(24352..24917)  
/note="match: GSS: Em:AQ667419"  
repeat\_region 24486..24656  
/note="MER20 repeat: matches 31..218 of consensus"  
repeat\_region 24804..24865  
/note="31 copies 2 mer ca 72% conserved"  
repeat\_region 25351..25559  
/note="MIR repeat: matches 32..249 of consensus"  
repeat\_region 25564..25754  
/note="MIR repeat: matches 77..262 of consensus"  
repeat\_region 25803..26008  
/note="MER58A repeat: matches 9..224 of consensus"  
repeat\_region 26250..26541  
/note="AlusX repeat: matches 1..301 of consensus"  
repeat\_region 26544..27216  
/note="L1MA5 repeat: matches 5469..6149 of consensus"  
repeat\_region 27217..27523  
Query Match 87.6%; Score 18.4; DB 9; Length 143981;  
Best Local Similarity 95.0%; Pred No.1.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20  
||||| |||||||  
Db 82141 ATGTGGGTGGTGGAAATAT 82160

RESULT 2  
AL662786/c  
LOCUS  
DEFINITION Mus musculus chromosome 11 clone RP23-340010, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL662786  
VERSION AL662786.2 GI:17903997  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
McLay, K.  
Direct Submission  
Submitted (03-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Dec 18, 2001 this sequence version replaced gi:17736632.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BM340010  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 202716 bases at least Q40  
Consensus quality: 203013 bases at least Q30  
Consensus quality: 203237 bases at least Q20  
Insert size: 203467; sum-of-contigs  
Insert size: 198131; 4.7% error; agarose-fp  
Quality coverage: 10.99x in Q20 bases; sum-of-contigs Quality  
coverage: 11.59x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
Location/Qualifiers  
source 1..204067  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-340010"  
/clone\_lib="RPCI-23"  
1..9424  
/note="assembly\_fragment:02278  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left"  
misc\_feature 9525..31155  
/note="assembly\_fragment:00677  
fragment\_chain:1"  
31256..89487  
/note="assembly\_fragment:01888  
fragment\_chain:1"  
89588..119319  
/note="assembly\_fragment:04012  
fragment\_chain:1"  
119420..175677  
/note="assembly\_fragment:03857  
fragment\_chain:1"  
175778..197522  
/note="assembly\_fragment:02822  
fragment\_chain:1"  
197623..204067  
/note="assembly\_fragment:03165  
fragment\_chain:1  
clone\_end:Y7  
vector\_side:right"  
BASE COUNT 57719 a 40407 c 41655 g 63676 t 610 others  
ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 204067;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatatc 21  
||||| |||||||  
Db 94801 ATGTGGGAGGTGGAAATATC 94781

RESULT 3  
HSA336359/c  
LOCUS HSA336359 737 bp DNA linear PRI 01-OCT-2001  
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone  
HSJ-CB3RS.  
ACCESSION AJ336359  
VERSION AJ336359.1 GI:15880777  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 737)  
AUTHORS Kutsenko, A. S., Gizatullin, R. Z., Al-Amin, A. N., Wang, F.,  
Podowski, R. M., Matushkin, Y. G., Kvasha, S. M., Gyanchandani, A.,  
Muravenko, O. V., Protokopov, A. I., Kashuba, V. I., Kisselev, L. L.,

Thu Jul 25 08:38:03 2002

Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
Analysis of NotI flanking sequences: a new tool for gene discovery  
and verification of the human genome

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 737)  
Zabarovsky, E.R.  
Direct Submission  
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden

FEATURES  
source  
1..737  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HSJ-CB3RS"  
138 a 229 c 224 g 146 t

BASE COUNT  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 737;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttctcgccgggtggaataatc 21  
|||||  
Db 328 TGTGGGGCGGGCAATATC 309

RESULT 4  
HSA336191/c  
LOCUS  
DEFINITION  
Homo sapiens genomic sequence surrounding NotI site, clone  
HSJ-B021RS.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
human.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 750)  
Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F.,  
Podowski, R.M., Matushkin, Y.G., Kvasha, S.M., Gyanchandani, A.,  
Muravenko, O.V., Protopopov, A.I., Kashuba, V.I., Kisselev, L.L.,  
Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
Analysis of NotI flanking sequences: a new tool for gene discovery  
and verification of the human genome  
Unpublished  
2 (bases 1 to 750)  
Zabarovsky, E.R.  
Direct Submission  
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden

FEATURES  
source  
1..750  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HSJ-B021RS"  
134 a 228 c 229 g 154 t 5 others

BASE COUNT  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 750;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttctcgccgggtggaataatc 21  
|||||  
Db 327 TGTGGGGCGGGCAATATC 308

RESULT 5  
HSA334149/c  
LOCUS  
DEFINITION  
Homo sapiens genomic sequence surrounding NotI site, clone  
HSJ-DL7RS.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
human.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 792)  
Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F.,  
Podowski, R.M., Matushkin, Y.G., Kvasha, S.M., Gyanchandani, A.,  
Muravenko, O.V., Protopopov, A.I., Kashuba, V.I., Kisselev, L.L.,  
Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
Analysis of NotI flanking sequences: a new tool for gene discovery  
and verification of the human genome  
Unpublished  
2 (bases 1 to 792)  
Zabarovsky, E.R.  
Direct Submission  
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden

FEATURES  
source  
1..792  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HSJ-DL7RS"  
147 a 235 c 242 g 165 t 3 others

BASE COUNT  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 792;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttctcgccgggtggaataatc 21  
|||||  
Db 327 TGTGGGGCGGGCAATATC 308

RESULT 6  
AC106623/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-224B2, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 46 unordered pieces.  
AC106623  
AC106623.1 GI:18139147  
HTG; HTGS\_PHASE1.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 88941)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,  
Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.N., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,



Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jacobson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
Loulseg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
Ogulu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,  
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,  
Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 88941)  
Worley, K.C.  
Direct Submission  
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GLSB  
Center clone name: CH230-224B2  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 64203 bases at least Q40  
Consensus quality: 70064 bases at least Q30  
Consensus quality: 75160 bases at least Q20  
Estimated insert size: 56201; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 46 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 3507: contig of 3507 bp in length  
\* 3508 3607: gap of unknown length  
\* 3608 7003: contig of 3396 bp in length  
\* 7103 7103: gap of unknown length  
\* 7104 10062: contig of 2959 bp in length  
\* 10063 10162: gap of unknown length  
\* 10163 13524: contig of 3362 bp in length  
\* 13525 13625: gap of unknown length  
\* 13625 16718: contig of 3094 bp in length  
\* 16719 16818: gap of unknown length  
\* 16819 20154: contig of 3336 bp in length  
\* 20155 20254: gap of unknown length  
\* 20255 22778: contig of 2524 bp in length  
\* 22779 22878: gap of unknown length

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

22879 25169: contig of 2291 bp in length  
25170 25269: gap of unknown length  
25270 27417: contig of 2148 bp in length  
27418 27517: gap of unknown length  
27518 29500: contig of 1983 bp in length  
29501 29601: gap of unknown length  
29601 31479: contig of 1879 bp in length  
31480 31579: gap of unknown length  
31580 34809: contig of 3230 bp in length  
34810 34909: gap of unknown length  
34910 36936: contig of 2027 bp in length  
36937 37036: gap of unknown length  
37037 38629: contig of 1593 bp in length  
38630 38729: gap of unknown length  
38730 40371: contig of 1642 bp in length  
40372 40471: gap of unknown length  
40472 41983: contig of 1512 bp in length  
41984 42083: gap of unknown length  
42084 44236: contig of 2153 bp in length  
44237 44336: gap of unknown length  
44337 46111: contig of 1775 bp in length  
46112 46211: gap of unknown length  
46212 47884: contig of 1673 bp in length  
47885 47984: gap of unknown length  
47985 50053: contig of 2069 bp in length  
50054 50153: gap of unknown length  
50154 51358: contig of 1205 bp in length  
51359 51458: gap of unknown length  
51459 53537: contig of 2079 bp in length  
53538 53637: gap of unknown length  
53638 55058: contig of 1421 bp in length  
55059 55158: gap of unknown length  
55159 56396: contig of 1238 bp in length  
56397 56496: gap of unknown length  
56497 57898: contig of 1402 bp in length  
57899 57998: gap of unknown length  
57999 59678: contig of 1880 bp in length  
59679 59778: gap of unknown length  
59779 60800: contig of 1022 bp in length  
60801 60900: gap of unknown length  
60901 61958: contig of 1058 bp in length  
61959 62058: gap of unknown length  
62059 63787: contig of 1729 bp in length  
63788 63887: gap of unknown length  
63888 65032: contig of 1145 bp in length  
65033 65132: gap of unknown length  
65133 67086: contig of 1954 bp in length  
67087 67186: gap of unknown length  
67187 69261: contig of 2075 bp in length  
69262 69361: gap of unknown length  
69362 70708: contig of 1347 bp in length  
70709 70808: gap of unknown length  
70809 71876: contig of 1068 bp in length  
71877 71976: gap of unknown length  
71977 73010: contig of 1034 bp in length  
73011 73110: gap of unknown length  
73111 74254: contig of 1144 bp in length  
74255 74354: gap of unknown length  
74355 75787: contig of 1433 bp in length  
75788 75887: gap of unknown length  
75888 77293: contig of 1406 bp in length  
77294 77393: gap of unknown length  
77394 78438: contig of 1045 bp in length  
78439 78538: gap of unknown length  
78539 79665: contig of 1127 bp in length  
79666 79765: gap of unknown length  
79766 81448: contig of 1683 bp in length  
81449 81548: gap of unknown length  
81549 82890: contig of 1342 bp in length  
82891 84611: contig of 1621 bp in length  
84612 84711: gap of unknown length  
84712 86141: contig of 1430 bp in length

Thu Jul 25 08:38:03 2002

```

* 86142 86241: gap of unknown length
* 86242 87699: contig of 1458 bp in length
* 87700 87799: gap of unknown length
* 87800 88941: contig of 1142 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..88941
                /organism="Rattus norvegicus"
                /db_xref="taxon:10116"
                /clone="CH230-224B2"
BASE COUNT 23763 a 19418 c 16127 g 25100 t 4533 others
ORIGIN

Query Match      80.0%; Score 16.8; DB 2; Length 88941;
Best Local Similarity 90.0%; Pred. No. 8.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaataatc 21
||||| ||||||| ||
Db 19329 TGTCGGGTGGGAAATCTC 19310

RESULT 7
AC021089/c
LOCUS
DEFINITION Homo sapiens chromosome 16 clone CTD-2264D9, WORKING DRAFT
SEQUENCE, 11 ordered pieces.
ACCESSION AC021089
VERSION AC021089.5 GI:13786358
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100912)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100912)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:9964760.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 729006
Center clone name: CITB-H1_2264D9
-----
Summary Statistics
Consensus quality: 87446 bases at least Q40
Consensus quality: 96488 bases at least Q30
Consensus quality: 98216 bases at least Q20
Estimated insert size: 106000; agarose-fp estimation
Estimated insert size: 100412; sum-of-ctigs estimation
Quality coverage: 6.03 in Q20 bases; agarose-fp estimation
Quality coverage: 6.37 in Q20 bases; sum-of-ctigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 4525: contig of 4525 bp in length
* 4526 4625: gap of unknown length
* 4626 21040: contig of 16415 bp in length

```

```

* 21041 21140: gap of unknown length
* 21141 24617: contig of 3477 bp in length
* 24618 27291: gap of unknown length
* 24718 27291: contig of 2574 bp in length
* 27292 36976: gap of unknown length
* 27391 36976: contig of 9585 bp in length
* 36977 47599: gap of unknown length
* 37077 47599: contig of 10523 bp in length
* 47600 47699: gap of unknown length
* 47700 68271: contig of 20572 bp in length
* 68272 80965: contig of 12594 bp in length
* 80966 81066: gap of unknown length
* 81066 87307: contig of 6242 bp in length
* 87308 87408: gap of unknown length
* 87408 97702: contig of 10294 bp in length
* 97702 97802: gap of unknown length
* 97802 100912: contig of 3111 bp in length.
FEATURES
    Location/Qualifiers
        1..100912
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="CTD-2264D9"
            /clone_lib="Caltech human BAC library D"
BASE COUNT 27774 a 20918 c 22009 g 29159 t 1052 others
ORIGIN

Query Match      80.0%; Score 16.8; DB 2; Length 100912;
Best Local Similarity 90.0%; Pred. No. 8.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaataatc 21
||||| ||||||| ||
Db 84084 TCTCGGCTGTGGAATATC 84065

RESULT 8
AP004257/c
LOCUS
DEFINITION Oryza sativa chromosome 2 clone OJ1369_G08, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AP004257
VERSION AP004257.1 GI:16215756
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 104188)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
clone:OJ1369_G08
JOURNAL Published Only in Database (2001) In press
REFERENCE 2 (bases 1 to 104188)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.

```

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

source  
 1. 104188  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /chromosome="2"  
 /clone="QJ1369.G08"  
 BASE COUNT 30752 a 21170 c 21376 g 30890 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 2; Length 104188;  
 Best Local Similarity 90.0%; Pred. No. 8.1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20  
 ||||| ||||| |||||  
 Db 54563 ATGTCGGCGGCAGAAATAT 54544

## RESULT 9

AL139235\_3

## WPCOMMENT

Sequence split into 4 fragments LOCUS AL139235 Accession AL139235

Fragment Name	Begin	End
AL139235_0	1	110000
AL139235_1	100001	210000
AL139235_2	200001	310000
AL139235_3	300001	470025

Continuation (4 of 4) of AL139235 from base 300001 (AL139235 Homo sapiens chromosome 9 c

## Query Match

Best Local Similarity 80.0%; Score 16.8; DB 2; Length 107025;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgtcggcggtggaatatc 21  
 ||||| ||||| |||||

Db 16779 TCTCGGCGTGTGAATATC 16798

## RESULT 10

AC106570/c

## LOCUS

AC106570 172300 bp DNA linear HTG 12-JAN-2002  
 Rattus norvegicus clone CH230-119J12, \*\*\* SEQUENCING IN PROGRESS

## DEFINITION

\*\*\*, 67 unordered pieces.

## ACCESSION

AC106570.1 GI:18139094

## VERSION

HTG; HTGS\_PHASE1.

## KEYWORDS

Norway rat.

## SOURCE

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 172300)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,  
 Benton,J., Blincoe,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowie,S., Brice,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,F., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,  
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Lounsbury,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
 Ogihara,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Pery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,K., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,  
 Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,  
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Wasczyk,R., Wooden,S.,  
 Watlington,S., Williams,G., Williamson,A., Wlaczky,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 172300)

Worley,K.C.

Direct Submission

## JOURNAL

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

\* 41577 41676: gap of unknown length  
 \* 41677 45269: contig of 3593 bp in length  
 \* 45270 45369: gap of unknown length  
 \* 45370 48710: contig of 3341 bp in length  
 \* 48711 48810: gap of unknown length  
 \* 48811 51891: contig of 3071 bp in length  
 \* 51892 51981: gap of unknown length  
 \* 51982 55149: contig of 3168 bp in length  
 \* 55150 55249: gap of unknown length  
 \* 55250 58785: contig of 3536 bp in length  
 \* 58786 58885: gap of unknown length  
 \* 58886 62064: contig of 3179 bp in length  
 \* 62065 62164: gap of unknown length  
 \* 62165 66443: contig of 4279 bp in length  
 \* 66444 66543: gap of unknown length  
 \* 66544 68545: contig of 2002 bp in length  
 \* 68546 68645: gap of unknown length  
 \* 68646 71555: contig of 2910 bp in length  
 \* 71556 71656: gap of unknown length  
 \* 71657 74370: contig of 2715 bp in length  
 \* 74371 74470: gap of unknown length  
 \* 74471 77892: contig of 3422 bp in length  
 \* 77893 77992: gap of unknown length  
 \* 77993 79589: contig of 1597 bp in length  
 \* 79590 79689: gap of unknown length  
 \* 79690 83383: contig of 3694 bp in length  
 \* 83384 83483: gap of unknown length  
 \* 83484 87482: contig of 3999 bp in length  
 \* 87483 87582: gap of unknown length  
 \* 87583 90181: contig of 2599 bp in length  
 \* 90182 90281: gap of unknown length  
 \* 90282 92948: contig of 2667 bp in length  
 \* 92949 93048: gap of unknown length  
 \* 93049 95578: contig of 2530 bp in length  
 \* 95579 95678: gap of unknown length  
 \* 95679 98931: contig of 3253 bp in length  
 \* 98932 99031: gap of unknown length  
 \* 99032 101875: contig of 2844 bp in length  
 \* 101876 101975: gap of unknown length  
 \* 101976 104567: contig of 2592 bp in length  
 \* 104568 104667: gap of unknown length  
 \* 104668 106694: contig of 2026 bp in length  
 \* 106694 106794: gap of unknown length  
 \* 106794 108615: contig of 1821 bp in length  
 \* 108615 108715: gap of unknown length  
 \* 108715 110578: contig of 1864 bp in length  
 \* 110579 110678: gap of unknown length  
 \* 110679 114080: contig of 3402 bp in length  
 \* 114081 114180: gap of unknown length  
 \* 114181 115694: contig of 1514 bp in length  
 \* 115695 115794: gap of unknown length  
 \* 115795 118273: contig of 2479 bp in length  
 \* 118274 118373: gap of unknown length  
 \* 118374 120549: contig of 2176 bp in length  
 \* 120550 120649: gap of unknown length  
 \* 120650 123129: contig of 2480 bp in length  
 \* 123130 123229: gap of unknown length  
 \* 123230 124524: contig of 1295 bp in length  
 \* 124525 124624: gap of unknown length  
 \* 124625 126224: contig of 1600 bp in length  
 \* 126225 126324: gap of unknown length  
 \* 126325 128276: contig of 1952 bp in length  
 \* 128277 128376: gap of unknown length  
 \* 128377 129584: contig of 1208 bp in length  
 \* 129585 129684: gap of unknown length  
 \* 129685 130778: contig of 1094 bp in length  
 \* 130779 130879: gap of unknown length  
 \* 130879 132539: contig of 1661 bp in length  
 \* 132540 132640: gap of unknown length  
 \* 132640 134459: contig of 1819 bp in length  
 \* 134459 134558: gap of unknown length  
 \* 134559 135930: contig of 1372 bp in length  
 \* 135931 136030: gap of unknown length

\* 136031 137410: contig of 1380 bp in length  
 \* 137411 137510: gap of unknown length  
 \* 137511 138990: contig of 1480 bp in length  
 \* 138991 139090: gap of unknown length  
 \* 139091 140213: contig of 1123 bp in length  
 \* 140214 140313: gap of unknown length  
 \* 140314 141504: contig of 1191 bp in length  
 \* 141505 141604: gap of unknown length  
 \* 141605 142653: contig of 1049 bp in length  
 \* 142654 142753: gap of unknown length  
 \* 142754 144381: contig of 1628 bp in length  
 \* 144382 144881: gap of unknown length  
 \* 144882 145860: contig of 1379 bp in length  
 \* 145861 145960: gap of unknown length  
 \* 145961 147700: contig of 1740 bp in length  
 \* 147701 147800: gap of unknown length  
 \* 147801 149225: contig of 1425 bp in length  
 \* 149226 149325: gap of unknown length  
 \* 149326 151081: contig of 1756 bp in length  
 \* 151082 151181: gap of unknown length  
 \* 151182 152327: contig of 1146 bp in length  
 \* 152328 152427: gap of unknown length  
 \* 152428 154047: contig of 1620 bp in length  
 \* 154048 154147: gap of unknown length  
 \* 154148 155883: contig of 1736 bp in length  
 \* 155884 155983: gap of unknown length  
 \* 155984 157464: contig of 1481 bp in length

Query Match 80.0%; Score 16.8; DB 2; Length 172300;  
 Best Local Similarity 90.0%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20  
 |||||  
 Db 94051 ATGTAGGAGGTGGAAATAT 94032

RESULT 11

AC024791  
 LOCUS 194322 bp DNA linear INV 16-NOV-2001  
 DEFINITION Caenorhabditis elegans cosmid v47G6A, complete sequence.  
 AC024791  
 ACCESSION AC024791.1 GI:7140347  
 VERSION HTG.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 194322)  
 AUTHORS The C. elegans Sequencing Consortium.  
 TITLE Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 REFERENCE 2 (bases 1 to 194322)  
 AUTHORS Minx, P., Graves, T. and Hawrysko, C.  
 TITLE The sequence of C. elegans cosmid v47G6A  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 194322)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 194322)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 5 (bases 1 to 194322)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

6 (bases 1 to 194322)  
Waterston, R.  
Direct Submission  
Submitted (06-APR-2001)  
University, 4444 Fores  
7 (bases 1 to 194322)

7 (bases 1 to 194322)  
Waterston, R.

Submitted (30-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

8 (bases 1 to 194322)  
Waterston. R.

Submitted (05-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

LOUIS, MO 63110, USA  
9 (bases 1 to 194322)  
Waterston, R.

Direct Submission  
Submitted (16-NOV-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 444 Forest Park Avenue, St.  
Louis, MO 63110, USA

COMMENT

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RO, England  
email: [rwnemato@wustl.edu](mailto:rwnemato@wustl.edu) and [jcs@sanger.ac.uk](mailto:jcs@sanger.ac.uk)

**NOTICE:** This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

## NEIGHBORING COSMID INFORMATION

The YAC segment Y47G6A was sequenced to span the gap between K03B5 and T28F2. The clone sequenced to the left is K03B5, 200 bp overlap; the clone sequenced to the right is T28F2, 200 bp overlap. Actual start of this YAC is at base position 28973 of CELY44E3A; actual end is at 20751 of CELT28F2.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

	FEATURES	SOURCE
1.	100% Cotton	100%
2.	100% Cotton	100%
3.	100% Cotton	100%
4.	100% Cotton	100%
5.	100% Cotton	100%
6.	100% Cotton	100%
7.	100% Cotton	100%
8.	100% Cotton	100%
9.	100% Cotton	100%
10.	100% Cotton	100%
11.	100% Cotton	100%
12.	100% Cotton	100%
13.	100% Cotton	100%
14.	100% Cotton	100%
15.	100% Cotton	100%
16.	100% Cotton	100%
17.	100% Cotton	100%
18.	100% Cotton	100%
19.	100% Cotton	100%
20.	100% Cotton	100%
21.	100% Cotton	100%
22.	100% Cotton	100%
23.	100% Cotton	100%
24.	100% Cotton	100%
25.	100% Cotton	100%
26.	100% Cotton	100%
27.	100% Cotton	100%
28.	100% Cotton	100%
29.	100% Cotton	100%
30.	100% Cotton	100%
31.	100% Cotton	100%
32.	100% Cotton	100%
33.	100% Cotton	100%
34.	100% Cotton	100%
35.	100% Cotton	100%
36.	100% Cotton	100%
37.	100% Cotton	100%
38.	100% Cotton	100%
39.	100% Cotton	100%
40.	100% Cotton	100%
41.	100% Cotton	100%
42.	100% Cotton	100%
43.	100% Cotton	100%
44.	100% Cotton	100%
45.	100% Cotton	100%
46.	100% Cotton	100%
47.	100% Cotton	100%
48.	100% Cotton	100%
49.	100% Cotton	100%
50.	100% Cotton	100%
51.	100% Cotton	100%
52.	100% Cotton	100%
53.	100% Cotton	100%
54.	100% Cotton	100%
55.	100% Cotton	100%
56.	100% Cotton	100%
57.	100% Cotton	100%
58.	100% Cotton	100%
59.	100% Cotton	100%
60.	100% Cotton	100%
61.	100% Cotton	100%
62.	100% Cotton	100%
63.	100% Cotton	100%
64.	100% Cotton	100%
65.	100% Cotton	100%
66.	100% Cotton	100%
67.	100% Cotton	100%
68.	100% Cotton	100%
69.	100% Cotton	100%
70.	100% Cotton	100%
71.	100% Cotton	100%
72.	100% Cotton	100%
73.	100% Cotton	100%
74.	100% Cotton	100%
75.	100% Cotton	100%
76.	100% Cotton	100%
77.	100% Cotton	100%
78.	100% Cotton	100%
79.	100% Cotton	100%
80.	100% Cotton	100%
81.	100% Cotton	100%
82.	100% Cotton	100%
83.	100% Cotton	100%
84.	100% Cotton	100%
85.	100% Cotton	100%
86.	100% Cotton	100%
87.	100% Cotton	100%
88.	100% Cotton	100%
89.	100% Cotton	100%
90.	100% Cotton	100%
91.	100% Cotton	100%
92.	100% Cotton	100%
93.	100% Cotton	100%
94.	100% Cotton	100%
95.	100% Cotton	100%
96.	100% Cotton	100%
97.	100% Cotton	100%
98.	100% Cotton	100%
99.	100% Cotton	100%
100.	100% Cotton	100%

```

Location/Qualifiers
1. .194322
/organism="Caenorhabd
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="Y47G6A"
/complement(706. .5939
/gene="Y47G6A.12"
.note="sep-1"

```

/note=-sep-1.  
complement(join(706...1369,1422...1517,1568...1746,  
2101...2995,5046...3347,3399...3521,3571...3796,4529...4928,  
4986...5145,5196...5393))  
/gene=Y47G6A.12)  
/note=contains similarity to cell division-associated  
proteins; Emicella nidulans BMB (GB:M83232) and  
Schizosaccharomyces pombe CUT1 (GB:M36179); coded for by  
the following C. elegans cDNAs: yk40C5.3, yk40C5.5,  
yk40C5.6, yk40C5.7, yk40C5.8, yk40C5.9, yk40C5.10,  
yk40C5.11, yk40C5.12, yk40C5.13, yk40C5.14, yk40C5.15,  
yk40C5.16, yk40C5.17, yk40C5.18, yk40C5.19, yk40C5.20,  
yk40C5.21, yk40C5.22, yk40C5.23, yk40C5.24, yk40C5.25,  
yk40C5.26, yk40C5.27, yk40C5.28, yk40C5.29, yk40C5.30,  
yk40C5.31, yk40C5.32, yk40C5.33, yk40C5.34, yk40C5.35,  
yk40C5.36, yk40C5.37, yk40C5.38, yk40C5.39, yk40C5.40,  
yk40C5.41, yk40C5.42, yk40C5.43, yk40C5.44, yk40C5.45,  
yk40C5.46, yk40C5.47, yk40C5.48, yk40C5.49, yk40C5.50,  
yk40C5.51, yk40C5.52, yk40C5.53, yk40C5.54, yk40C5.55,  
yk40C5.56, yk40C5.57, yk40C5.58, yk40C5.59, yk40C5.60,  
yk40C5.61, yk40C5.62, yk40C5.63, yk40C5.64, yk40C5.65,  
yk40C5.66, yk40C5.67, yk40C5.68, yk40C5.69, yk40C5.70,  
yk40C5.71, yk40C5.72, yk40C5.73, yk40C5.74, yk40C5.75,  
yk40C5.76, yk40C5.77, yk40C5.78, yk40C5.79, yk40C5.80,  
yk40C5.81, yk40C5.82, yk40C5.83, yk40C5.84, yk40C5.85,  
yk40C5.86, yk40C5.87, yk40C5.88, yk40C5.89, yk40C5.90,  
yk40C5.91, yk40C5.92, yk40C5.93, yk40C5.94, yk40C5.95,  
yk40C5.96, yk40C5.97, yk40C5.98, yk40C5.99, yk40C5.100,  
yk40C5.101, yk40C5.102, yk40C5.103, yk40C5.104, yk40C5.105,  
yk40C5.106, yk40C5.107, yk40C5.108, yk40C5.109, yk40C5.110,  
yk40C5.111, yk40C5.112, yk40C5.113, yk40C5.114, yk40C5.115,  
yk40C5.116, yk40C5.117, yk40C5.118, yk40C5.119, yk40C5.120,  
yk40C5.121, yk40C5.122, yk40C5.123, yk40C5.124, yk40C5.125,  
yk40C5.126, yk40C5.127, yk40C5.128, yk40C5.129, yk40C5.130,  
yk40C5.131, yk40C5.132, yk40C5.133, yk40C5.134, yk40C5.135,  
yk40C5.136, yk40C5.137, yk40C5.138, yk40C5.139, yk40C5.140,  
yk40C5.141, yk40C5.142, yk40C5.143, yk40C5.144, yk40C5.145,  
yk40C5.146, yk40C5.147, yk40C5.148, yk40C5.149, yk40C5.150,  
yk40C5.151, yk40C5.152, yk40C5.153, yk40C5.154, yk40C5.155,  
yk40C5.156, yk40C5.157, yk40C5.158, yk40C5.159, yk40C5.160,  
yk40C5.161, yk40C5.162, yk40C5.163, yk40C5.164, yk40C5.165,  
yk40C5.166, yk40C5.167, yk40C5.168, yk40C5.169, yk40C5.170,  
yk40C5.171, yk40C5.172, yk40C5.173, yk40C5.174, yk40C5.175,  
yk40C5.176, yk40C5.177, yk40C5.178, yk40C5.179, yk40C5.180,  
yk40C5.181, yk40C5.182, yk40C5.183, yk40C5.184, yk40C5.185,  
yk40C5.186, yk40C5.187, yk40C5.188, yk40C5.189, yk40C5.190,  
yk40C5.191, yk40C5.192, yk40C5.193, yk40C5.194, yk40C5.195,  
yk40C5.196, yk40C5.197, yk40C5.198, yk40C5.199, yk40C5.200,  
yk40C5.201, yk40C5.202, yk40C5.203, yk40C5.204, yk40C5.205,  
yk40C5.206, yk40C5.207, yk40C5.208, yk40C5.209, yk40C5.210,  
yk40C5.211, yk40C5.212, yk40C5.213, yk40C5.214, yk40C5.215,  
yk40C5.216, yk40C5.217, yk40C5.218, yk40C5.219, yk40C5.220,  
yk40C5.221, yk40C5.222, yk40C5.223, yk40C5.224, yk40C5.225,  
yk40C5.226, yk40C5.227, yk40C5.228, yk40C5.229, yk40C5.230,  
yk40C5.231, yk40C5.232, yk40C5.233, yk40C5.234, yk40C5.235,  
yk40C5.236, yk40C5.237, yk40C5.238, yk40C5.239, yk40C5.240,  
yk40C5.241, yk40C5.242, yk40C5.243, yk40C5.244, yk40C5.245,  
yk40C5.246, yk40C5.247, yk40C5.248, yk40C5.249, yk40C5.250,  
yk40C5.251, yk40C5.252, yk40C5.253, yk40C5.254, yk40C5.255,  
yk40C5.256, yk40C5.257, yk40C5.258, yk40C5.259, yk40C5.260,  
yk40C5.261, yk40C5.262, yk40C5.263, yk40C5.264, yk40C5.265,  
yk40C5.266, yk40C5.267, yk40C5.268, yk40C5.269, yk40C5.270,  
yk40C5.271, yk40C5.272, yk40C5.273, yk40C5.274, yk40C5.275,  
yk40C5.276, yk40C5.277, yk40C5.278, yk40C5.279, yk40C5.280,  
yk40C5.281, yk40C5.282, yk40C5.283, yk40C5.284, yk40C5.285,  
yk40C5.286, yk40C5.287, yk40C5.288, yk40C5.289, yk40C5.290,  
yk40C5.291, yk40C5.292, yk40C5.293, yk40C5.294, yk40C5.295,  
yk40C5.296, yk40C5.297, yk40C5.298, yk40C5.299, yk40C5.300,  
yk40C5.301, yk40C5.302, yk40C5.303, yk40C5.304, yk40C5.305,  
yk40C5.306, yk40C5.307, yk40C5.308, yk40C5.309, yk40C5.310,  
yk40C5.311, yk40C5.312, yk40C5.313, yk40C5.314, yk40C5.315,  
yk40C5.316, yk40C5.317, yk40C5.318, yk40C5.319, yk40C5.320,  
yk40C5.321, yk40C5.322, yk40C5.323, yk40C5.324, yk40C5.325,  
yk40C5.326, yk40C5.327, yk40C5.328, yk40C5.329, yk40C5.330,  
yk40C5.331, yk40C5.332, yk40C5.333, yk40C5.334, yk40C5.335,  
yk40C5.336, yk40C5.337, yk40C5.338, yk40C5.339, yk40C5.340,  
yk40C5.341, yk40C5.342, yk40C5.343, yk40C5.344, yk40C5.345,  
yk40C5.346, yk40C5.347, yk40C5.348, yk40C5.349, yk40C5.350,  
yk40C5.351, yk40C5.352, yk40C5.353, yk40C5.354, yk40C5.355,  
yk40C5.356, yk40C5.357, yk40C5.358, yk40C5.359, yk40C5.

gene

```

/codon_start=1
/product="Hypothetical protein Y47G6A.13"
/protein_id="ME50672.1"

```

CDs

```

/db_xref="GI:7331985"
/translation="MTNSDQKTSSSASKVMRPGHIIDKRIKALCKGACGVVYKVED
KNRGYCAAMKEVDFSEDFDLQLEVNVLSYKLSQSDRVDKLIYDCKRKYLYRWVMTL
CGKDLMAIMKIQRENDATMTAMRVAYLVLGUKLQHEAGYVHVDKPGNIMTAAKGR
DARVALLDGMARTFVAGADGGKKLPMRRRIKPLGIRYCSVMHFERKEGRCDD
LVAMITYSTVLTGLPNKDEKDEIMSKSKTSLDLPIDELKSTFYETLRTLSY
ADRNPKYKLYNLLMAKLRKLNFLPQWEDENDERMAKAEKKEKKEKKEKKEK
KENELSDMSMTKTIAPKAKASAEVAGEDEOKVKYDMSSEGOODTLPADKGNKLIS

```

gene

CDS

```
6514. .15110
/gene="Y47G6A.28"
join(6514. .6585,6645. .6750,11554. .12015,12506. .12599,
12648. .13282,14331. .14479,14814. .15110)
```

/note="coded for by the following C. elegans cDNAs: CEMSC37F, vk277a4.5, vk281a5.5, vk497f2.5, CEMSC37#

```
CEMSC3/E, YKZ1
/codon start=1
```

```
/codon_start=1
/product="Hypothetical protein Y47G6A.28"
/protein_id="AAK95891.1"
```

/translation: "MPRIKQAPT KHRSPKMRSEADSEFLDDEEERIKKEQLQRVN  
EKKKHIOQRALQESQALTSDESNEARSGSPARNLVEGFKVKERYRNIENHIE  
EYKAREGSECPSEVNEIIGNVYFPTMSAAQOYFVNRFWMTLKGSGVAPR  
NISISQFDNARALKQOLDQFRKLKNSYVNOVEDDYVJGNGRRCTITLSSHGVSIG  
LRRSPSPSSKSTAP LKSHYKRSFSESLDENVSPSPSRHSRVEDRPTTSSNAKPFES  
DDESEBEATDDESGSDDEVLDTDEDESVNSNLIRPKLIMEYTPSPKIOISLFT  
LSYWPGRNGIKLKEFPQIPAKLEISTPKPLTSSSLAEPEKENTKPNPSYSLRKRAY  
DESKKPKGSDSPVNSPTVPVTPVTRKRSKAPEDNGTEKQKQKAPYCNHCSLL  
KEKKAKLEAKNPISSTGTPVAVTQPKAKKSEKVTPKCOCSKQNRSTVAATEPKY  
KIADKSKVKSFFTKIKVLWLTITNFCCNGKGEVKNNAKTRFSNKKPKVFETKNIQFT  
KIADKVVKNSFFLSESDSDIISIPSSMAVADVTLELV"

gene

**CDS**

```
complement(join(6843. .6916,7020. .7100,7152. .7539,
7591. .7680,7952. .8438,9257. .9483,9535. 9594))
```

/gene="Y47G6A.l3"  
/note="contains similarity to Pfam family PF00069  
(Eukaryotic protein kinase domain), score=53.0, E=1.3e-13,  
N=1"

```
N=1
/codon start=1
```

```
/codon_start=1
/product="Hypothetical protein Y47G6A.13"
/protein_id="ME50672.1"
```

CDs

```

/db_xref="GI:7331985"
/translation="MTNSDQKTSSSASKVMRPGHIIDKRIKALCKGACGVVYKVED
KNRGYCAAMKEVDFSEDFDLQLEVNVLSYKLSQSDRVDKLIYDCKRKYLYRWMTVL
CGKDLMAIMKIQRENDATMTAMRVAYLVLGUKLQHEAGYVHVDKPGNIMTAAKGR
DARVALLDGMARTFVAGADGGKKLPMRRRIKPLGIRYCSVMHFKEKRGDCDD
LVAMITYSTVLTGLPNKDEKDEIMMSKSTIKLALFDPELKSFTYELRTLSY
ADRNPKYKLYNLLMAKYLKLNFLPQEWEDENMAKAEKBEKBEKBEKBEKBEK
KENELSDMSMTKTIAPKAKASBEVAGEDEOKVKYDMSSEGOODTLYADKGNKLIS

```

```
AKDMASRSEVNENQYGTAKAGNSVTGDSIDMETATECDQMEYNKKALPRDELQF
VYPAIAPRNFTEVKIPP"
gene
16733..29478
/genes="Y47G6A.11"
/note="msh-6"
join(16733..16868,17929..18209,19439..20203,22406..23211,
24513..24924,25923..26474,27901..28233,29203..29478)
CDS
24513..24924,25923..26474,27901..28233,29203..29478)
/genes="Y47G6A.11"
/note="contains similarity to Pfam family PF00488 (DNA
mismatch repair proteins, mutS family), score=296.7,
E=2.9e-85, N=1; coded for by the following C. elegans
cDNAs: yk1b11.3, yk1b11.5, yk8g10.3, yk8g10.5, yk44a10.5,
yk47b4.3, yk47b4.5, yk151b12.3, yk170g6.5, yk170g6.3,
yk290a8.3, yk290a8.5, yk321b5.3, yk321b5.5, yk563a9.5,
yk788e06.5, yk788e06.3, yk563a9.3"
/codon_start=1
/product="Hypothetical protein Y47G6A.11"
/protein_id="AAK95890.1"
/db_xref="GI:15375264"
/translation="MSKQSLMSFFTKPKSEKPEEVKESVVEPKSLKNDTPKIS

Query Match      80.0%; Score 16.8; DB 3; Length 194322;
Best Local Similarity 90.0%; Pred. No. 7.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgtcggcggtgaaatat 20
|||||
Db 68177 ATGTGGCGCTGGAACTAT 68196

RESULT 12
AC006782/c 319857 bp DNA linear HTG 25-FEB-1999
LOCUS
DEFINITION
Caenorhabditis elegans clone Y47G6, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
AC006782
AC006782.2 GI:4309789
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
Caenorhabditis elegans.
SOURCE
Caenorhabditis elegans.
ORGANISM
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 319857)
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 319857)
Waterston,R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263156.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3652: contig of 3652 bp in length
3653 3666: gap of unknown length
3667 36798: contig of 33132 bp in length
36799 36812: gap of unknown length
36813 156230: contig of 119418 bp in length
156231 156244: gap of unknown length
156245 319857: contig of 163613 bp in length.
Location/Qualifiers
1. 319857
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y47G6"

FEATURES
source
1
662: contig of 662 bp in length
663 762: gap of 100 bp
763 1465: contig of 703 bp in length
```

\* 1466 1565: gap of 100 bp  
\* 1566 2258: contig of 693 bp in length  
\* 2259 2358: gap of 100 bp  
\* 2359 3036: contig of 678 bp in length  
\* 3037 3136: gap of 100 bp  
\* 3137 3826: contig of 690 bp in length  
\* 3827 3926: gap of 100 bp  
\* 3927 4615: contig of 689 bp in length  
\* 4616 4715: gap of 100 bp  
\* 4716 5405: contig of 690 bp in length  
\* 5406 5505: gap of 100 bp  
\* 5506 6194: contig of 689 bp in length  
\* 6195 6294: gap of 100 bp  
\* 6295 7005: contig of 711 bp in length  
\* 7006 7105: gap of 100 bp  
\* 7106 7803: contig of 698 bp in length  
\* 7804 7903: gap of 100 bp  
\* 7904 8593: contig of 690 bp in length  
\* 8594 8693: gap of 100 bp  
\* 8694 9393: contig of 700 bp in length  
\* 9394 9493: gap of 100 bp  
\* 9494 10189: contig of 696 bp in length  
\* 10190 10289: gap of 100 bp  
\* 10290 10985: contig of 696 bp in length  
\* 10986 11085: gap of 100 bp  
\* 11086 11772: contig of 687 bp in length  
\* 11773 11872: gap of 100 bp  
\* 11873 12564: contig of 692 bp in length  
\* 12565 12664: gap of 100 bp  
\* 12665 13371: contig of 707 bp in length  
\* 13372 13471: gap of 100 bp  
\* 13472 14172: contig of 701 bp in length  
\* 14173 14272: gap of 100 bp  
\* 14273 14959: contig of 687 bp in length  
\* 14960 15059: gap of 100 bp  
\* 15060 15741: contig of 682 bp in length  
\* 15742 15841: gap of 100 bp  
\* 15842 16533: contig of 692 bp in length  
\* 16534 16633: gap of 100 bp  
\* 16634 17326: contig of 693 bp in length  
\* 17327 17426: gap of 100 bp  
\* 17427 18121: contig of 695 bp in length  
\* 18122 18221: gap of 100 bp  
\* 18222 18932: contig of 711 bp in length  
\* 18933 19032: gap of 100 bp  
\* 19033 19734: contig of 702 bp in length  
\* 19735 19834: gap of 100 bp  
\* 19835 20517: contig of 683 bp in length  
\* 20518 20617: gap of 100 bp  
\* 20618 21314: contig of 697 bp in length  
\* 21315 21414: gap of 100 bp  
\* 21415 22122: contig of 708 bp in length  
\* 22123 22222: gap of 100 bp  
\* 22223 22912: contig of 690 bp in length  
\* 22913 23012: gap of 100 bp  
\* 23013 23710: contig of 698 bp in length  
\* 23711 23810: gap of 100 bp  
\* 23811 24509: contig of 699 bp in length  
\* 24510 24609: gap of 100 bp  
\* 24610 25298: contig of 689 bp in length  
\* 25299 25398: gap of 100 bp  
\* 25399 26091: contig of 693 bp in length  
\* 26092 26191: gap of 100 bp  
\* 26192 26895: contig of 704 bp in length  
\* 26896 26995: gap of 100 bp  
\* 26996 27687: contig of 692 bp in length  
\* 27688 27787: gap of 100 bp  
\* 27788 28407: contig of 620 bp in length  
\* 28408 28507: gap of 100 bp  
\* 28508 29209: contig of 702 bp in length  
\* 29210 29309: gap of 100 bp  
\* 29310 29996: contig of 687 bp in length  
\* 29997 30096: gap of 100 bp

\* 30097 30805: contig of 709 bp in length  
\* 30806 30905: gap of 100 bp  
\* 30906 31594: contig of 689 bp in length  
\* 31595 31694: gap of 100 bp  
\* 31695 32380: contig of 686 bp in length  
\* 32381 32480: gap of 100 bp  
\* 32481 33170: contig of 690 bp in length  
\* 33171 33270: gap of 100 bp  
\* 33271 33970: contig of 700 bp in length  
\* 33971 34070: gap of 100 bp  
\* 34071 34757: contig of 687 bp in length  
\* 34758 34857: gap of 100 bp  
\* 34858 35545: contig of 688 bp in length  
\* 35546 35645: gap of 100 bp  
\* 35646 36325: contig of 680 bp in length  
\* 36326 36425: gap of 100 bp  
\* 36426 37126: contig of 701 bp in length  
\* 37127 37226: gap of 100 bp  
\* 37227 37920: contig of 694 bp in length  
\* 37921 38020: gap of 100 bp  
\* 38021 38711: contig of 691 bp in length  
\* 38712 38811: gap of 100 bp  
\* 38812 39501: contig of 690 bp in length  
\* 39502 39601: gap of 100 bp  
\* 39602 40281: contig of 680 bp in length  
\* 40282 40381: gap of 100 bp  
\* 40382 41083: contig of 702 bp in length  
\* 41084 41183: gap of 100 bp  
\* 41184 41881: contig of 698 bp in length  
\* 41882 41981: gap of 100 bp  
\* 41982 42681: contig of 700 bp in length  
\* 42682 42781: gap of 100 bp  
\* 42782 43464: contig of 683 bp in length  
\* 43465 43564: gap of 100 bp  
\* 43565 44243: contig of 679 bp in length  
\* 44244 44343: gap of 100 bp  
\* 44344 45036: contig of 693 bp in length  
\* 45037 45136: gap of 100 bp  
\* 45137 45822: contig of 686 bp in length  
\* 45823 45922: gap of 100 bp  
\* 45923 46641: contig of 719 bp in length  
\* 46642 46741: gap of 100 bp  
\* 46742 47433: contig of 692 bp in length  
\* 47434 47533: gap of 100 bp  
\* 47534 48228: contig of 695 bp in length  
\* 48229 48328: gap of 100 bp  
\* 48329 49006: contig of 678 bp in length  
\* 49007 49106: gap of 100 bp  
\* 49107 49796: contig of 690 bp in length  
\* 49797 49896: gap of 100 bp  
\* 49897 50589: contig of 693 bp in length  
\* 50590 50689: gap of 100 bp  
\* 50690 51377: contig of 688 bp in length  
\* 51378 51477: gap of 100 bp  
\* 51478 52172: contig of 695 bp in length  
\* 52173 52272: gap of 100 bp  
\* 52273 52987: contig of 715 bp in length  
\* 52988 53087: gap of 100 bp  
\* 53088 53803: contig of 716 bp in length  
\* 53804 53903: gap of 100 bp  
\* 53904 54593: contig of 690 bp in length  
\* 54594 54693: gap of 100 bp  
\* 54694 55397: contig of 704 bp in length  
\* 55398 55497: gap of 100 bp  
\* 55498 56204: contig of 707 bp in length  
\* 56205 56304: gap of 100 bp  
\* 56305 57013: contig of 709 bp in length.

FEATURES Location/Qualifiers  
1. .57013

Query Match 78.1%; Score 16.4; DB 2; Length 57013;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	tcgggcgttggaataac 21                         TCGGGGGTTGGAATAC 28548
Db	28531	
RESULT	14	
AC0911196		63528 bp DNA linear HTG 03-APR-2001
LOCUS		Homo sapiens chromosome 11 clone RP11-371C18 map 11, LOW-PASS
DEFINITION		SEQUENCE SAMPLING.
AC0911196		AC0911196
VERSION		AC0911196.1 GI:13518192
KEYWORDS		HTG; HTGS_PHASE0.
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 63528)
AUTHORS		Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE		Homo sapiens chromosome 11, clone RP11-371C18
REFERENCE		Unpublished
AUTHORS		2 (bases 1 to 63528) Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckgalter,B., Brown,A., Cammarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K., Lamarez,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., MacDonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE		Direct Submission
JOURNAL		Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT		All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WITBR Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a> ----- Project Information Center project name: LI3097 Center clone name: 371_C_18 ----- * NOTE: This record contains 77 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for * identifying clones that may be gene-rich and allows * overlap relationships among clones to be deduced. * However, it should not be assumed that this clone * will be sequenced to completion. In the event that * the record is updated, the accession number will * be preserved. 723: contig of 723 bp in length * 1 * 724 823: gap of 100 bp * 824 1583: contig of 760 bp in length



\* 31455 32147: contig of 693 bp in length  
\* 32148 32247: gap of 100 bp  
\* 32248 32975: contig of 728 bp in length  
\* 32976 33075: gap of 100 bp  
\* 33076 33804: contig of 729 bp in length  
\* 33805 33904: gap of 100 bp  
\* 33905 34665: contig of 761 bp in length  
\* 34666 34765: gap of 100 bp  
\* 34766 35467: contig of 702 bp in length  
\* 35468 35567: gap of 100 bp  
\* 35568 36316: contig of 749 bp in length  
\* 36317 36416: gap of 100 bp  
\* 36417 37129: contig of 713 bp in length  
\* 37130 37229: gap of 100 bp  
\* 37230 37953: contig of 724 bp in length  
\* 37954 38053: gap of 100 bp  
\* 38054 38763: contig of 710 bp in length  
\* 38764 38863: gap of 100 bp  
\* 38864 39582: contig of 719 bp in length  
\* 39583 39682: gap of 100 bp  
\* 39683 40424: contig of 742 bp in length  
\* 40425 40524: gap of 100 bp  
\* 40525 41236: contig of 712 bp in length  
\* 41237 41336: gap of 100 bp  
\* 41337 42037: contig of 701 bp in length  
\* 42038 42137: gap of 100 bp  
\* 42138 42869: contig of 732 bp in length  
\* 42870 42969: gap of 100 bp  
\* 42970 43660: contig of 691 bp in length  
\* 43661 43760: gap of 100 bp  
\* 43761 44460: contig of 700 bp in length  
\* 44461 44560: gap of 100 bp  
\* 44561 45293: contig of 733 bp in length  
\* 45294 45393: gap of 100 bp  
\* 45394 46142: contig of 749 bp in length  
\* 46143 46242: gap of 100 bp  
\* 46243 46941: contig of 699 bp in length  
\* 46942 47041: gap of 100 bp  
\* 47042 47821: contig of 780 bp in length  
\* 47822 47921: gap of 100 bp  
\* 47922 48704: contig of 783 bp in length  
\* 48705 48804: gap of 100 bp  
\* 48805 49564: contig of 760 bp in length  
\* 49565 49664: gap of 100 bp  
\* 49665 50364: contig of 700 bp in length  
\* 50365 50464: gap of 100 bp  
\* 50465 51179: contig of 715 bp in length  
\* 51180 51279: gap of 100 bp  
\* 51280 52001: contig of 722 bp in length  
\* 52002 52101: gap of 100 bp  
\* 52102 52822: contig of 721 bp in length  
\* 52823 52922: gap of 100 bp  
\* 52923 53650: contig of 728 bp in length  
\* 53651 53750: gap of 100 bp  
\* 53751 54459: contig of 709 bp in length  
\* 54460 54559: gap of 100 bp  
\* 54560 55296: contig of 737 bp in length  
\* 55297 55396: gap of 100 bp  
\* 55397 56105: contig of 709 bp in length  
\* 56106 56205: gap of 100 bp  
\* 56206 56968: contig of 763 bp in length  
\* 56969 57068: gap of 100 bp

Query Match 78.1%; Score 16.4; DB 2; Length 63528;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tcggcggtggaatc 21  
||||| |||||||  
DB 34789 TCGGGGTGGAATATC 34806

RESULT 15

AC010774  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS

AC010774 100511 bp DNA linear HTG 13-JUL-2000  
Homo sapiens clone RP11-2G12, LOW-PASS SEQUENCE SAMPLING.  
AC010774  
AC010774.3 GI:9119885  
HTG: HTGS\_PHASE0.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 100511)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-2G12  
Unpublished  
2 (bases 1 to 100511)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,A., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:5910818.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L2531  
Center clone name: 2\_G\_12  
-----

\* NOTE: This record contains 105 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 846: contig of 846 bp in length  
\* 847 946: gap of 100 bp  
\* 947 1762: contig of 816 bp in length  
\* 1763 1862: gap of 100 bp  
\* 1863 2715: contig of 853 bp in length  
\* 2716 2815: gap of 100 bp  
\* 2816 3678: contig of 863 bp in length  
\* 3679 3778: gap of 100 bp  
\* 3779 4598: contig of 820 bp in length  
\* 4599 4698: gap of 100 bp  
\* 4699 5538: contig of 840 bp in length  
\* 5539 5638: gap of 100 bp  
\* 5639 6483: contig of 845 bp in length  
\* 6484 6583: gap of 100 bp  
\* 6584 7400: contig of 817 bp in length  
\* 7401 7500: gap of 100 bp  
\* 7501 8349: contig of 849 bp in length

\* 8350 8449: gap of 100 bp  
\* 8450 9306: contig of 857 bp in length  
\* 9307 9406: gap of 100 bp  
\* 9407 10280: contig of 874 bp in length  
\* 10281 10380: gap of 100 bp  
\* 10381 11235: contig of 855 bp in length  
\* 11236 11335: gap of 100 bp  
\* 11336 12194: contig of 859 bp in length  
\* 12195 12294: gap of 100 bp  
\* 12295 13133: contig of 839 bp in length  
\* 13134 13233: gap of 100 bp  
\* 13234 14096: contig of 863 bp in length  
\* 14097 14196: gap of 100 bp  
\* 14197 15064: contig of 868 bp in length  
\* 15065 15164: gap of 100 bp  
\* 15165 16038: contig of 874 bp in length  
\* 16039 16138: gap of 100 bp  
\* 16139 16989: contig of 851 bp in length  
\* 16990 17089: gap of 100 bp  
\* 17090 17928: contig of 839 bp in length  
\* 17929 18028: gap of 100 bp  
\* 18029 18849: contig of 821 bp in length  
\* 18850 18949: gap of 100 bp  
\* 18950 19797: contig of 848 bp in length  
\* 19798 19897: gap of 100 bp  
\* 19898 20763: contig of 866 bp in length  
\* 20764 20863: gap of 100 bp  
\* 20864 21761: contig of 898 bp in length  
\* 21762 21861: gap of 100 bp  
\* 21862 22725: contig of 864 bp in length  
\* 22726 22825: gap of 100 bp  
\* 22826 23705: contig of 880 bp in length  
\* 23706 23805: gap of 100 bp  
\* 23806 24664: contig of 859 bp in length  
\* 24665 24764: gap of 100 bp  
\* 24765 25614: contig of 850 bp in length  
\* 25615 25714: gap of 100 bp  
\* 25715 26555: contig of 841 bp in length  
\* 26556 26655: gap of 100 bp  
\* 26656 27508: contig of 853 bp in length  
\* 27509 27608: gap of 100 bp  
\* 27609 28472: contig of 864 bp in length  
\* 28473 28572: gap of 100 bp  
\* 28573 29439: contig of 867 bp in length  
\* 29440 29539: gap of 100 bp  
\* 29540 30406: contig of 867 bp in length  
\* 30407 30506: gap of 100 bp  
\* 30507 31382: contig of 876 bp in length  
\* 31383 31482: gap of 100 bp  
\* 31483 32342: contig of 860 bp in length  
\* 32343 32442: gap of 100 bp  
\* 32443 33319: contig of 877 bp in length  
\* 33320 33419: gap of 100 bp  
\* 33420 34276: contig of 857 bp in length  
\* 34277 34376: gap of 100 bp  
\* 34377 35258: contig of 882 bp in length  
\* 35259 35358: gap of 100 bp  
\* 35359 36218: contig of 860 bp in length  
\* 36219 36318: gap of 100 bp  
\* 36319 37161: contig of 843 bp in length  
\* 37162 37261: gap of 100 bp  
\* 37262 38153: contig of 892 bp in length  
\* 38154 38253: gap of 100 bp  
\* 38254 39076: contig of 823 bp in length  
\* 39077 39176: gap of 100 bp  
\* 39177 40044: contig of 868 bp in length  
\* 40045 40144: gap of 100 bp  
\* 40145 41010: contig of 866 bp in length  
\* 41011 41110: gap of 100 bp  
\* 41111 42014: contig of 904 bp in length  
\* 42015 42114: gap of 100 bp  
\* 42115 42982: contig of 868 bp in length  
\* 42983 43082: gap of 100 bp

\* 43083 43950: contig of 868 bp in length  
\* 43951 44050: gap of 100 bp  
\* 44051 44917: contig of 867 bp in length  
\* 44918 45017: gap of 100 bp  
\* 45018 45895: contig of 878 bp in length  
\* 45896 45995: gap of 100 bp  
\* 45996 46867: contig of 872 bp in length  
\* 46868 46967: gap of 100 bp  
\* 46968 47866: contig of 899 bp in length  
\* 47867 47966: gap of 100 bp  
\* 47967 48832: contig of 866 bp in length  
\* 48833 48932: gap of 100 bp  
\* 48933 49799: contig of 867 bp in length  
\* 49800 49899: gap of 100 bp  
\* 49900 50753: contig of 854 bp in length  
\* 50754 50853: gap of 100 bp  
\* 50854 51718: contig of 865 bp in length  
\* 51719 51818: gap of 100 bp  
\* 51819 52705: contig of 887 bp in length  
\* 52706 52805: gap of 100 bp  
\* 52806 53655: contig of 850 bp in length  
\* 53656 53755: gap of 100 bp  
\* 53756 54592: contig of 837 bp in length  
\* 54593 54692: gap of 100 bp  
\* 54693 55567: contig of 875 bp in length  
\* 55568 55667: gap of 100 bp  
\* 55668 56545: contig of 878 bp in length  
\* 56546 56645: gap of 100 bp  
\* 56646 57481: contig of 836 bp in length  
\* 57482 57581: gap of 100 bp  
\* 57582 58427: contig of 846 bp in length  
\* 58428 58527: gap of 100 bp  
\* 58528 59378: contig of 851 bp in length  
\* 59379 59478: gap of 100 bp  
\* 59479 60348: contig of 870 bp in length  
\* 60349 60448: gap of 100 bp  
\* 60449 61307: contig of 859 bp in length  
\* 61308 61407: gap of 100 bp  
\* 61408 62279: contig of 872 bp in length  
\* 62280 62379: gap of 100 bp  
\* 62380 63208: contig of 829 bp in length  
\* 63209 63308: gap of 100 bp  
\* 63309 64163: contig of 855 bp in length  
\* 64164 64263: gap of 100 bp  
\* 64264 65109: contig of 846 bp in length  
\* 65110 65209: gap of 100 bp  
\* 65210 66079: contig of 870 bp in length  
\* 66080 66179: gap of 100 bp  
\* 66180 67031: contig of 852 bp in length  
\* 67032 67131: gap of 100 bp  
\* 67132 67974: contig of 843 bp in length  
\* 67975 68074: gap of 100 bp  
\* 68075 68904: contig of 830 bp in length  
\* 68905 69004: gap of 100 bp  
\* 69005 69857: contig of 853 bp in length

Query Match 78.18; Score 16.4; DB 2: Length 100511;  
Best Local Similarity 94.48; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgcggcggtggaaata 19  
| | | | | | | | | | | | | | | | | | | | |  
Db 47835 TGTCGGCGGTGGAAAAA 47852

Search completed: July 25, 2002, 05:38:07  
Job time: 6640 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:44:40 ; Search time 378.07 Seconds  
(without alignments)  
95.366 Million cell updates/sec

Title: US-09-235-416-3

Perfect score: 21  
Sequence: 1 atgtcggcggtggaatc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_032802:*			
1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*		
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*		
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*		
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*		
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*		
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*		
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*		
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*		
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*		
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*		
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*		
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*		
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*		
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*		
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*		
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*		
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*		
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*		
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*		
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*		
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*		
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*		
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*		
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	100.0	21	20	AAH87657	Kinesin motor prot
2	21	100.0	2352	20	AAH87656	Thermomyces lanugi
C 3	16.2	77.1	2707	22	AAH44925	CDNA encoding nove
C 4	15.8	75.2	152	21	AAH45638	Human secreted exp
5	15.8	75.2	704	21	AAA01667	Human colon cancer
6	15.8	75.2	1209	21	AAA64407	Open reading frame
7	15.8	75.2	1449	21	AAA62003	Hydrophobic domain
8	15.8	75.2	1652	21	AAA62013	Hydrophobic domain
9	15.8	75.2	1793	22	AHH14868	Human CDNA sequenc

10	15.8	75.2	2252	21	AAA64406	DNA encoding a hum
11	15.8	75.2	2285	22	AAH87657	Human CDNA sequenc
12	15.8	75.2	2437	22	AAH83002	Human MBSP10 polyp
13	15.8	75.2	2449	22	AAH59608	Human cell cycle a
14	15.8	75.2	2495	22	AAH45041	CDNA encoding nove
15	15.8	75.2	2505	21	AAH76426	Human ORFX ORF1981
16	15.8	75.2	2505	21	AAH23447	CDNA encoding nove
17	15.8	75.2	2607	22	AAH45229	Human ASB-3 protei
18	15.8	75.2	4623	22	ABA09048	Human immune syste
19	15.8	75.2	12393	24	ABL33263	Human ASB-3 protei
20	15.4	73.3	8035	21	AAH51305	A. terreus LovF/Sc
21	15.4	73.3	31328	21	AAH51295	A. terreus DAB/lov
c	23	15.2	60	20	AAH19638	Complement system
c	24	15.2	61	20	AAH19678	Complement system
25	15.2	72.4	442	21	AAH00237	Human secreted pro
26	15.2	72.4	503	21	AAH77917	CDNA encoding huma
27	15.2	72.4	503	22	AAH28655	Colon tumour relat
28	15.2	72.4	503	22	AAH29078	Colon tumour relat
29	15.2	72.4	780	23	AAH56003	Salmonella typhi D
30	15.2	72.4	1319	21	AAH77873	Human cancer assoc
c	31	15.2	1425	23	AAH76017	DNA encoding novel
c	32	15.2	1933	22	AAH85566	Human immune/haema
33	15.2	72.4	2700	22	AAH74536	A human alpha-2 ma
34	15.2	72.4	2739	22	AAH74534	A human alpha-2 ma
c	35	15.2	3227	23	ABL20897	Drosophila melanog
c	36	15.2	3239	23	ABL02077	Drosophila melanog
37	15.2	72.4	3405	22	AAH74533	A human alpha-2 ma
38	15.2	72.4	3413	22	ABA14721	Human nervous syst
39	15.2	72.4	4065	22	AAH74535	DNA encoding a hum
40	15.2	72.4	4527	22	AAH74529	A human alpha-2 ma
41	15.2	72.4	4527	22	AAH74530	A human alpha-2 ma
c	42	15.2	4734	23	ABL13666	Drosophila melanog
43	15.2	72.4	4771	22	AAH74532	A human alpha-2 ma
44	15.2	72.4	4771	23	AAH5411	DNA encoding novel
c	45	15.2	5343	23	ABL20896	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAH87657	
ID	AAH87657 standard; DNA; 21 BP.
XX	
AC	AAH87657;
DT	26-OCT-1999 (first entry)
XX	
DE	Kinesin motor protein TL-gamma gene PCR primer.
XX	
KW	TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;
KW	neurodegenerative disease; Alzheimer's disease;
KW	Parkinson's disease; Huntington's disease;
KW	anyotrophic lateral sclerosis; PCR; primer; ss.
XX	
OS	Synthetic.
OS	Thermomyces lanuginosus.
XX	
PN	WO9937659-A1.
XX	
PD	29-JUL-1999.
XX	
PF	22-JAN-1999; 99WO-US01355.
PR	23-JAN-1998; 98US-0072361.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Goldstein LSB, Sakowicz R;
XX	
DR	WPI; 1999-493950/41.
XX	

PT New nucleic acid encoding microtubule motor protein, used for  
PT diagnosis of fungal infection and neurodegenerative disease  
XX  
XX  
PS Claim 7; Page 60; 75pp; English.  
XX  
CC This oligonucleotide represents one of a claimed primer pair (see  
CC also AAX87658) that can be used in the amplification of microtubule  
CC motor protein nucleic acids, especially TL-gamma DNA or RNA from a  
CC hyphal fungus such as Thermomyces lanuginosus (see AAX87656).  
CC TL-gamma (see AAX06618) is a novel ATP-dependent, plus end-directed  
CC microtubule motor protein that is a member of the unc-104 family  
CC and kinesin superfamily. The invention provides TL-gamma nucleic  
CC acids, proteins and antibodies, and methods of screening for  
CC TL-gamma modulators potentially useful for treating hyphal fungal  
CC infections and diseases caused by mutated TL-gamma, e.g.  
CC neurodegeneration. TL-gamma nucleic acids are also used to identify  
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or  
CC other motor proteins, to generate protein structural models, for  
CC recombinant production of TL-gamma as antisense molecules, to  
CC produce transgenic or knockout animals, and in gene therapy.  
CC Detection of TL-gamma allows differentiation between hyphal and  
CC non-hyphal fungal infections.  
XX  
SQ Sequence 21 BP; 5 A; 3 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.24; 0; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaataatc 21  
|||||  
Db 1 atgtcggcggtggaataatc 21

RESULT 2  
AAX87656  
ID AAX87656 standard; DNA; 2352 BP.  
XX  
AC AAX87656;  
XX  
DT 26-OCT-1999 (first entry)  
XX  
DE Thermomyces lanuginosus kinesin motor protein TL-gamma DNA.  
XX  
XX TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;  
KW neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease;  
KW anyotrophic lateral sclerosis; ss.  
XX  
XX Thermomyces lanuginosus.  
OS  
XX  
XX WO9937659-A1.  
PN  
XX 29-JUL-1999.  
PD  
XX 22-JAN-1999; 99WO-US01355.  
PF  
XX 23-JAN-1998; 98US-0072361.  
PR  
XX (REGC ) UNIV CALIFORNIA.  
PA  
XX Goldstein LSB, Sakowicz R;  
PI  
XX WPI: 1999-493950/41.  
XX P-PSDB; AAY06618.  
DR  
XX New nucleic acid encoding microtubule motor protein, used for  
PT diagnosis of fungal infection and neurodegenerative disease  
XX  
XX Claim 4; Page 69-70; 75pp; English.  
XX  
XX This is the DNA sequence of the Thermomyces lanuginosus gene

CC coding for TL-gamma (see AAY06618), a novel ATP-dependent, plus  
CC end-directed microtubule motor protein that is a member of the  
CC unc-104 family and kinesin superfamily. The gene was isolated  
CC from genomic DNA by PCR amplification (see also AAX87659-61).  
CC The invention provides TL-gamma nucleic acids, proteins and  
CC antibodies, and methods of screening for TL-gamma modulators  
CC potentially useful for treating hyphal fungal infections and  
CC diseases caused by mutated TL-gamma, e.g. neurodegeneration  
CC involving anterograde axonal transport, such as Alzheimer's,  
CC Parkinson's or Huntington's diseases or amyotrophic lateral  
CC sclerosis. TL-gamma nucleic acids are also used to identify  
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or  
CC other motor proteins, to generate protein structural models, for  
CC recombinant production of TL-gamma, as antisense molecules, to  
CC produce transgenic or knockout animals, and in gene therapy.  
CC Detection of TL-gamma allows differentiation between hyphal and  
CC non-hyphal fungal infections.  
XX  
SQ Sequence 2352 BP; 626 A; 564 C; 672 G; 490 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2352;  
Best Local Similarity 100.0%; Pred. No. 0.34;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaataatc 21  
|||||  
Db 1 atgtcggcggtggaataatc 21

RESULT 3  
AAS44925/c  
ID AAS44925 standard; cDNA; 2707 BP.  
XX  
AC AAS44925;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human secretory protein, Seq ID No 6.  
XX  
XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
KW anyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
KW fertility; analgesic; pain; antigen; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200166689-A2.  
PN  
XX 13-SEP-2001.  
PD  
XX 05-MAR-2001; 2001WO-US04942.  
PF  
XX 07-MAR-2000; 2000US-0519705.  
PR 19-MAY-2000; 2000US-0574454.  
PR 17-JUN-2000; 2000US-0596193.  
PR 14-JUL-2000; 2000US-0616847.  
PR 19-SEP-2000; 2000US-0665363.  
PR 20-OCT-2000; 2000US-0693267.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
XX WPI: 2001-589934/66.  
XX P-PSDB; AAU28025.  
DR  
XX



KW Human; colon cancer; tumour; diagnosis; gene expression product;  
 KW probe; detection; cancerous state; metastasis; identification;  
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09958675-A2.  
 XX  
 XX 18-NOV-1999.  
 XX  
 XX 13-MAY-1999; 99WO-US10602.  
 XX  
 PR 14-MAY-1998; 98US-0085426.  
 PR 15-MAY-1998; 98US-0085537.  
 PR 15-MAY-1998; 98US-0085596.  
 PR 21-OCT-1998; 98US-0105234.  
 PR 27-OCT-1998; 98US-0105877.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 XX  
 XX WPI: 2000-126369/11.  
 DR  
 XX Polynucleotide library used to determine cancerous states of mammalian  
 PT cells -  
 XX  
 PS Claim 1; Page 599-600; 1097pp; English.  
 XX  
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
 CC libraries constructed from human colon cancer cell lines. The present  
 CC invention also describes a method of detecting differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell, comprising  
 CC detecting at least one differentially expressed gene product in a test  
 CC sample derived from a cell suspected of being cancerous, where detection  
 CC of the differentially expressed gene product is correlated with a  
 CC cancerous state of the cell from which the test sample was derived.  
 CC The polynucleotide sequences can be used in a method for detecting  
 CC differentially expressed genes correlated with a cancerous state of a  
 CC mammalian cell. The polynucleotides can also be used as probes for  
 CC detecting and mapping related genes. They can be used in diagnosis and  
 CC prognosis of diseases and disorders (e.g. identification of  
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
 CC responsiveness of cancer to therapy). This is particularly for breast  
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
 CC negative breast cancer, lung cancer, and colon cancer.  
 XX  
 SQ Sequence 704 BP; 217 A; 134 C; 153 G; 194 T; 6 other;  
 Query Match 75.2%; Score 15.8; DB 21; Length 704;  
 Best Local Similarity 89.5%; Pred. NO. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 tgcggcggtggaatat 20  
 ||||| |||||  
 DB 214 tgcgggtggtggaatat 232  
 RESULT 6  
 AAA64407  
 ID AAA64407 standard; cDNA; 1209 BP.  
 XX  
 AC AAA64407;  
 XX  
 XX 02-JAN-2001 (first entry)  
 DT  
 XX Open reading frame of DNA encoding a human TANGO 201 polypeptide.

XX TANGO 201; TANGO 223; secreted protein; transmembrane protein;  
 KW proliferative disorder; neoplasm; tumour; pancreatic disorder; cirrhosis;  
 KW pancreatitis; thyroiditis; goiter; Graves' disease; gastric disorder;  
 KW gastritis; tumour; placental disorder; placentalitis; spontaneous abortion;  
 KW pulmonary disorder; atelectasis; edema; Goodpasture's syndrome;  
 KW muscular dystrophy; cardiovascular disorder; ischemic heart disease;  
 KW congenital heart disease; cerebral edema; cerebrovascular disease;  
 KW hepatic disorder; jaundice; hepatitis; leukocytic disorder; hyperplasia;  
 KW leukopenia; leukocytosis; malignant lymphoma; prostate disorder;  
 KW inflammatory disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1209  
 FT /\*tag= a "TANGO 201"  
 FT /product= "TANGO 201"  
 FT sig\_peptide 1..99  
 FT /\*tag= b  
 FT mat\_peptide 100..1209  
 FT /\*tag= c  
 FT  
 XX W0200050442-A2.  
 PN  
 XX 31-AUG-2000.  
 PD  
 XX 25-FEB-2000; 2000WO-US04784.  
 PF  
 XX 26-FEB-1999; 99US-0259388.  
 PR  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX McCarthy SA;  
 PI  
 XX WPI: 2000-572066/53.  
 DR P-PSDB; AAB08640.  
 XX  
 PT Novel nucleic acid molecule encoding secreted or transmembrane protein  
 PT useful for identifying modulators and for diagnosing and treating  
 PT pancreatic, cardiovascular, liver and pituitary disorders -  
 XX  
 PS Claim 2; Page 120-121; 176pp; English.  
 CC The present sequence encodes a TANGO 201 polypeptide. The specification  
 CC also describes a TANGO 223 polypeptide. These polypeptides are secreted  
 CC or transmembrane proteins. Human TANGO 201 and 223 nucleic acids,  
 CC proteins and their modulators are useful for treating proliferative  
 CC disorders e.g. neoplasms or tumours, pancreatic disorders  
 CC (e.g. pancreatitis), disorders of the adrenal cortex, adrenal medulla,  
 CC thyroid gland (e.g. thyroiditis), goiter, Graves' disease, gastric  
 CC disorders (e.g. gastritis or tumours), placental disorders (e.g.  
 CC placentalitis or spontaneous abortion), pulmonary disorders (e.g.  
 CC atelectasis), edema, Goodpasture's syndrome, disorders of the skeletal  
 CC muscle (e.g. muscular dystrophy), cardiovascular disorders (e.g.  
 CC ischemic heart disease and congenital heart disease), disorders of the  
 CC brain (e.g. cerebral edema), cerebrovascular disease and to treat  
 CC injury or trauma to the brain. They are also useful for treating  
 CC hepatic disorders (e.g. jaundice, hepatitis, cirrhosis or malignant  
 CC tumours), renal, testicular, intestinal disorders. TANGO 223  
 CC polynucleotides are also useful for treating leukocytic disorders (e.g.  
 CC leukopenias, leukocytosis and malignant lymphomas) and prostate  
 CC disorders (e.g. inflammatory diseases, hyperplasia or tumours).  
 XX  
 SQ Sequence 1209 BP; 380 A; 247 C; 282 G; 300 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 1209;  
 Best Local Similarity 89.5%; Pred. NO. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 tgcggcggtggaatat 20  
 ||||| |||||

```

Db 882 tgcggttggtggaatat 900
RESULT 7
AAA62003
ID AAA62003 standard; DNA; 1449 BP.
XX
AC AAA62003;
XX
DT 02-FEB-2001 (first entry)
XX
DE Hydrophobic domain protein cDNA HP03091 isolated from Liver cells.
XX
KW Human; secreted protein; membrane protein; hydrophobic domain;
KW proliferation control; differentiation induction; material transport;
KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200029448-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-JP06412.
XX
PR 17-NOV-1998; 98JP-0326255.
XX
PR 22-DEC-1998; 98JP-0364315.
XX
PR 16-MAR-1999; 99JP-0069811.
XX
PR 27-APR-1999; 99JP-0119299.
XX
PR 19-MAY-1999; 99JP-0138169.
XX
PA (SAGA ) SAGAMI CHEM RES CENT.
XX
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
WPI: 2000-387753/33.
XX
P-PSDB; AAB12130.
XX
Proteins comprising hydrophobic regions, such as secretory and membrane
PT proteins, useful in research and diagnostics and having various
PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,
PT hemostatic, thrombolytic -
XX
Claim 3; Page 248-249; 410pp; English.
XX
Secretory proteins play important roles in the proliferation control, the
CC differentiation induction, the material transport and the biophylaxis of
CC cells. Membrane proteins have important roles as signal receptors, ion
CC channels and transporters. The present sequence is the coding sequence
CC for a human protein which has at least one hydrophobic domain. The
CC protein encoded by the present sequence may be a secretory or a membrane
CC protein. The encoded protein may have cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, anti-inflammatory activity and tumour
CC inhibition activity. The present sequence could therefore be used for
CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC disease, and cancer via gene therapy.
XX
Sequence 1449 BP; 458 A; 292 C; 333 G; 366 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 1449;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcggcggtggaatat 20

```

```

Db 1044 tgcggttggtggaatat 1062
RESULT 8
AAA62013
ID AAA62013 standard; DNA; 1652 BP.
XX
AC AAA62013;
XX
DT 02-FEB-2001 (first entry)
XX
DE Hydrophobic domain protein cDNA HP03091 isolated from Liver cells.
XX
KW Human; secreted protein; membrane protein; hydrophobic domain;
KW proliferation control; differentiation induction; material transport;
KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200029448-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-JP06412.
XX
PR 17-NOV-1998; 98JP-0326255.
XX
PR 22-DEC-1998; 98JP-0364315.
XX
PR 16-MAR-1999; 99JP-0069811.
XX
PR 27-APR-1999; 99JP-0119299.
XX
PR 19-MAY-1999; 99JP-0138169.
XX
PA (SAGA ) SAGAMI CHEM RES CENT.
XX
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
WPI: 2000-387753/33.
XX
P-PSDB; AAB12130.
XX
Proteins comprising hydrophobic regions, such as secretory and membrane
PT proteins, useful in research and diagnostics and having various
PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,
PT hemostatic, thrombolytic -
XX
Claim 4; Page 267-270; 410pp; English.
XX
Secretory proteins play important roles in the proliferation control, the
CC differentiation induction, the material transport and the biophylaxis of
CC cells. Membrane proteins have important roles as signal receptors, ion
CC channels and transporters. The present sequence is the coding sequence
CC for a human protein which has at least one hydrophobic domain. The
CC protein encoded by the present sequence may be a secretory or a membrane
CC protein. The encoded protein may have cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, anti-inflammatory activity and tumour
CC inhibition activity. The present sequence could therefore be used for
CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC disease, and cancer via gene therapy.
XX
Sequence 1652 BP; 526 A; 323 C; 379 G; 424 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 1652;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```





CC injury or trauma to the brain. They are also useful for treating  
 CC hepatic disorders (e.g. jaundice, hepatitis, cirrhosis or malignant  
 CC tumours), renal, testicular, intestinal disorders. TANGO 223  
 CC polynucleotides are also useful for treating leukocytic disorders (e.g.  
 CC leukopenias, leukocytosis and malignant lymphomas) and prostate  
 CC disorders (e.g. inflammatory diseases, hyperplasia or tumours).  
 XX Sequence 2252 BP; 681 A; 444 C; 496 G; 631 T; 0 other;  
 SQ

Query Match 75.2%; Score 15.8; DB 21; Length 2252;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaatat 20  
 Db 1060 tgctgggtggaatat 1078

RESULT 11  
 AAS21290  
 ID AAS21290 standard; cDNA; 2285 BP.  
 AC AAS21290;  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human cDNA sequence encoding for PRO4321 polypeptide.  
 XX  
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;  
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200140466-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 01-DEC-2000; 2000WO-US32678.  
 XX  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 09-DEC-1999; 99US-0170262.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30911.  
 PR 20-DEC-1999; 99WO-US30999.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US04914.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 21-MAR-2000; 2000WO-US07377.  
 PR 21-MAR-2000; 2000WO-US07532.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 10-NOV-2000; 2000WO-US30873.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 XX WPI; 2001-403281/43.  
 DR P-PSDB; AAU12218.  
 XX  
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect  
 PT other PRO polypeptides, link bioactive molecules to cells expressing  
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.  
 PT lung, breast, prostate, cervical  
 XX  
 PS Claim 3; Fig 93; 813pp; English.  
 XX  
 CC AAS21244-AAS21518 encode for novel human secretory and transmembrane  
 CC PRO polypeptides. The PRO polypeptides are useful to detect other  
 CC PRO polypeptides, to link bioactive molecules to cells expressing  
 CC PRO polypeptides, to modulate biological activities of cells expressing  
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
 CC polypeptide expression in a cell sample to that in a control sample.  
 CC Some of the 275 sequences are also useful to stimulate the release of  
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
 CC proliferation or differentiation of chondrocytes, the proliferation or  
 CC gene expression in pericyte cells, the release of proteoglycans from  
 CC cartilage, the proliferation of inner ear utricular supporting cells or  
 CC of T-lymphocytes, or the release of a cytokine from peripheral blood  
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify  
 CC molecules involved in binding interactions. The polynucleotides encoding  
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
 CC transgenic or knock out animals and can be used in gene therapy.  
 XX  
 SQ Sequence 2285 BP; 693 A; 442 C; 502 G; 648 T; 0 other;

Query Match 75.2%; Score 15.8; DB 22; Length 2285;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaatat 20  
 Db 1046 tgctgggtggaatat 1064

RESULT 12  
 AAF83002  
 ID AAF83002 standard; cDNA; 2437 BP.  
 XX  
 AC AAF83002;  
 XX  
 DT 29-JUN-2001 (first entry)  
 XX  
 DE Human MBSP10 polypeptide encoding cDNA (clone 20604798.0.1).  
 XX  
 KW MBSPX; cancer; preclampsia; immune system; neurological; cytostatic;  
 KW gynecological; antiinflammatory; neuroprotective; inotropic; relaxant;  
 KW cardiac; dermatological; gene therapy; human; MBSP10; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 147..1597  
 FT /\*tag= a  
 FT /product= "MBSP10"  
 XX  
 PN WO200127277-A2.  
 XX  
 PD 19-APR-2001.  
 XX  
 PF 13-OCT-2000; 2000WO-US28480.  
 XX

PR 13-OCT-1999; 99US-0159231.  
PR 12-JAN-2000; 2000US-0175570.  
PR 12-OCT-2000; 2000US-0159231.  
XX (CURA-) CURAGEN CORP.  
XX PA  
XX PA  
XX Shimkets RA, Lichenstein H, Boldog FL;  
PI  
XX  
DR WPI; 2001-282030/29.  
DR P-PSDB; AAB62406.  
XX  
XX Novel human polynucleotide sequences and the membrane bound or secreted  
PT polypeptides encoded by these sequences, designated MBSPX -  
XX  
PS Claim 9; Page 44-47; 157pp; English.  
XX  
XX The invention relates to novel polypeptides, termed MBSPX and  
CC polynucleotides encoding the MBSPX polypeptides. The MBSPX polypeptide,  
CC nucleic acid and an MBSPX antibody are useful for treating or preventing  
CC a pathology associated with the protein especially in humans. The MBSPX  
CC nucleic acid can be used to express MBSPX protein (e.g. via a recombinant  
CC expression vector in a host cell in gene therapy applications), an to  
CC detect MBSPX mRNA in a biological sample or a genetic lesion in a MBSPX  
CC gene. Disorders associated with insufficient or excessive production of  
CC MBSPX protein include cancer, preclampsia, immune system disorders and  
CC inflammation, neurological disorders, cardiovascular disorders; and skin  
CC and muscle abnormalities. The anti-MBSPX antibodies can be used to detect  
CC and isolate MBSPX proteins and modulate MBSPX activity. The present  
CC sequence represents the nucleotide sequence of MBSPX10.  
XX  
SQ Sequence 2437 BP; 739 A; 473 C; 530 G; 695 T; 0 other;  
  
Query Match 75.2%; Score 15.8; DB 22; Length 2437;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 tgctggcggtgggaaatat 20  
Db 1190 tgctgggtggtgggaaatat 1208  
||||| |||||||||  
  
RESULT 13  
AAF59608  
ID AAF59608 standard; cDNA; 2449 BP.  
XX  
XX AAF59608;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Human cell cycle and proliferation protein CCYPR-19 cDNA, SEQ ID NO:73.  
XX  
KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
KW antagonist; gene therapy; detection; gene therapy;  
KW transgenic animal disease model; immune disorder;  
KW developmental disorder; cell signalling disorder;  
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
KW menstrual cycle disorder; bacterial infection; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200107471-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000WO-US19948.  
XX  
XX 21-JUL-1999; 99US-0145075.  
PR 08-SEP-1999; 99US-0153129.  
PR 10-NOV-1999; 99US-0164647.  
XX  
XX (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
XX  
DR WPI; 2001-112727/12.  
DR P-PSDB; AAB60471.  
XX  
XX Human cell cycle and proliferation proteins and polynucleotides are  
PT used to treat, diagnose and prevent immune, developmental and cell  
PT signalling disorders and cell proliferative disorders including cancer -  
XX  
XX Example V; Page 179-180; 205pp; English.  
PS  
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
CC associated with decreased expression of functional CCYPR, while CCYPR  
CC antagonists are used to treat diseases or conditions associated with  
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
CC that specifically bind to CCYPR, and in drug screening methods to  
CC identify compounds that modulate the activity of CCYPR. CCYPR  
CC nucleotides can be used to generate transgenic animal models of human  
CC disease, and can be used in gene therapy in target cells with genetic  
CC abnormalities with respect to the expression of CCYPR for the  
CC treatment or prevention of a disorder associated with CCYPR.  
CC Diseases which can be diagnosed, treated and prevented using CCYPR  
CC proteins, nucleic acids, agonists or antagonists include immune,  
CC developmental and cell signalling disorders, and cell proliferative  
CC disorders including cancer. Specific examples of these disorders  
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
CC diabetes mellitus, disorders of the menstrual cycle and infections  
CC caused by bacteria.  
XX  
SQ Sequence 2449 BP; 749 A; 478 C; 531 G; 691 T; 0 other;  
  
Query Match 75.2%; Score 15.8; DB 22; Length 2449;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 tgctggcggtgggaaatat 20  
Db 1203 tgctgggttgggaaatat 1221  
||||| |||||||||  
  
RESULT 14  
AAS45041  
ID AAS45041 standard; cDNA; 2495 BP.  
XX  
XX AAS45041;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human secretory protein, Seq ID No 122.  
XX  
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
KW fertility; analgesic; pain; antigen; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200166689-A2.  
XX  
PD 13-SEP-2001.

XX PF 05-MAR-2001; 2001WO-US04942.  
 XX PR 07-MAR-2000; 2000US-0519705.  
 XX PR 19-MAY-2000; 2000US-0574454.  
 XX PR 17-JUN-2000; 2000US-0596193.  
 XX PR 14-JUL-2000; 2000US-0616847.  
 XX PR 19-SEP-2000; 2000US-0665363.  
 XX PR 20-OCT-2000; 2000US-0693267.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
 XX PI Zhao QA, Yang Y, Brmanac RT, Zhang J, Chen R, xue AJ, Wang J;  
 XX DR WPI; 2001-589934/66.  
 XX DR P-PSDB; AAU28141.  
 XX PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 XX PT prepared from various human tissues, for diagnosis and treatment of  
 XX PT cancer, neurological, inflammatory, and autoimmune disorders -  
 XX PS Claim 1; SEQ ID No 122; 107pp; English.  
 XX CC The invention relates to novel isolated human secreted polypeptides (I)  
 XX CC and polynucleotides (II). (I) and (II) are useful for treating  
 XX CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
 XX CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
 XX CC involved in increasing haematopoiesis, stem cell survival, bone growth  
 XX CC and remodeling. (I), (II) and modulators of (II) are useful for  
 XX CC prophylaxis or treatment of one or more cancers. (II) is also useful for  
 XX CC creating transgenic animals useful for studying the in vivo activities of  
 XX CC the polypeptide as well as for studying modulators of the polypeptides.  
 XX CC (I) induces the proliferation of neural cells and regeneration of nerve  
 XX CC and brain tissue and is useful for the treatment of central and  
 XX CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 XX CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 XX CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
 XX CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 XX CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 XX CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 XX CC tissue growth, and in tissue repair, healing of burns, incisions,  
 XX CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
 XX CC disorders, or periodontal disease. Furthermore, (I) is also useful for  
 XX CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 XX CC reperfusion injury in various tissues, various immune deficiencies and  
 XX CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 XX CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 XX CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 XX CC reactions and conditions, such as asthma or other respiratory problems.  
 XX CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
 XX CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 XX CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 XX CC analgesic effects or other pain reducing effects, immunoglobulin like  
 XX CC activity and can act as an antigen in a vaccine composition to raise an  
 XX CC immune response. AAS44920-AAS45295 represent novel human secreted protein  
 XX CC coding sequences of the invention.  
 XX SQ Sequence 2495 BP; 759 A; 484 C; 553 G; 699 T; 0 other;  
 Query Match 75.2%; Score 15.8; DB 22; Length 2495;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 tgctggcggtggaatat 20  
 ||||| ||||| |||||  
 Db 1236 tgctgggtggtggaatat 1254  
 RESULT 15  
 AAC76426  
 ID AAC76426 standard; cDNA; 2505 BP.

XX AC AAC76426;  
 XX DT 08-FEB-2001 (first entry)  
 XX DE Human ORFX ORF1981 polynucleotide sequence SEQ ID NO:3961.  
 XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW KW antiviral; antibacterial; antifungal; antirheumatic; antihyroid;  
 KW KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW KW thrombosis; contraceptive; ss.  
 XX OS Homo sapiens.  
 XX PN WO200058473-A2.  
 XX PD 05-OCT-2000.  
 XX PF 31-MAR-2000; 2000WO-US08621.  
 XX PR 31-MAR-1999; 99US-0127607.  
 XX PR 02-APR-1999; 99US-0127636.  
 XX PR 05-APR-1999; 99US-0127728.  
 XX PR 30-MAR-2000; 2000US-0540763.  
 XX PA (CURA-) CURAGEN CORP.  
 XX PI Shinkets RA, Leach M;  
 XX PI WPI; 2000-602362/57.  
 XX DR P-PSDB; AAB42217.  
 XX PT Novel nucleic acids and peptides derived from open reading frame X,  
 XX PT useful for treating e.g. cancers, proliferative disorders,  
 XX PT neurodegenerative disorders and cardiovascular disease -  
 XX PS Claim 5; Page 3123-3124; 5507pp; English.  
 XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 XX CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 XX CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 XX CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 XX CC antihyroid; and antianaemic. The sequences can be used for determining  
 XX CC the presence of or predisposition to, or preventing or treating  
 XX CC pathological conditions associated with an ORFX-associated disorder. The  
 XX CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 XX CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 XX CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 XX CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX SQ Sequence 2505 BP; 774 A; 482 C; 547 G; 700 T; 2 other;

Query Match 75.2%; Score 15.8; DB 21; Length 2505;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgtcgggcgtggaatat 20  
||||| |||||||||  
Db 1215 tgtcgttggtggaatat 1233

Search completed: July 25, 2002, 05:44:43  
Job time: 6951 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:38:29 ; Search time 90.93 Seconds  
(without alignments)  
56.728 Million cell updates/sec

Title: US-09-235-416-3  
Perfect score: 21  
Sequence: 1 atgtcggcggtggaataatc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.2	72.4	60	3	US-09-023-228B-91
C 2	15.2	72.4	61	3	US-09-023-228B-131
C 3	15.2	72.4	448	4	US-09-060-756-666
C 4	15.2	72.4	2424	1	US-08-821-119-16
C 5	15.2	72.4	3350	3	US-08-617-8608-3
6	15	71.4	1533	1	US-07-721-761A-32
7	15	71.4	1533	1	US-07-978-687-32
8	15	71.4	1533	5	PCT-US91-01746-12
9	15	71.4	1533	5	PCT-US91-05801-32
10	15	71.4	1533	1	US-08-616-368A-23
11	14.8	70.5	113	3	US-09-054-298-23
12	14.8	70.5	113	3	US-08-818-655-23
13	14.8	70.5	840	1	US-08-616-368A-14
14	14.8	70.5	840	3	US-09-054-298-14
15	14.8	70.5	840	4	US-08-818-655-14
16	14.6	69.5	1052	6	5489533-1
17	14.6	69.5	1052	6	5512660-1
C 18	14.2	67.6	769	4	US-08-776-971-118
C 19	14.2	67.6	769	4	US-08-776-971-119
C 20	14.2	67.6	1193	4	US-09-347-798-1
C 21	14.2	67.6	1193	4	US-09-347-798-1
22	14.2	67.6	1668	2	US-08-505-377-2
23	14.2	67.6	1668	2	US-08-505-377-5
24	14.2	67.6	1668	3	US-08-798-269-5
25	14.2	67.6	1668	3	US-08-798-269-5
26	14.2	67.6	1668	4	US-09-055-210-2
27	14.2	67.6	1668	4	US-09-055-210-5

28	14.2	67.6	1952	3	US-08-714-918-39	Sequence 39, Appl
29	14.2	67.6	1952	4	US-09-265-315-39	Sequence 39, Appl
30	14.2	67.6	1952	4	US-09-265-315-39	Sequence 39, Appl
31	14.2	67.6	1952	4	US-09-266-417-39	Sequence 39, Appl
C 32	14.2	67.6	2341	3	US-08-714-918-102	Sequence 102, App
C 33	14.2	67.6	2341	4	US-09-265-315-102	Sequence 102, App
C 34	14.2	67.6	2341	4	US-09-265-315-102	Sequence 102, App
C 35	14.2	67.6	2341	4	US-09-265-315-102	Sequence 102, App
36	14	66.7	935	3	US-09-187-331-3	Sequence 3, Appl1
37	14	66.7	935	4	US-09-470-946-3	Sequence 267, App
38	13.8	65.7	49	1	US-08-171-389-267	Sequence 267, App
39	13.8	65.7	49	1	US-08-123-936-267	Sequence 267, App
40	13.8	65.7	49	2	US-08-475-228A-267	Sequence 267, App
41	13.8	65.7	49	3	PCT-US93-12388-267	Sequence 267, App
42	13.8	65.7	827	4	US-08-976-259-36	Sequence 36, Appl
C 43	13.8	65.7	840	4	US-08-998-416-507	Sequence 507, App
C 44	13.8	65.7	1303	2	US-08-793-410-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-09-023-228B-91/c  
; Sequence 91, Application US/09023228B  
; Patent No. 6140490  
; GENERAL INFORMATION:  
; APPLICANT: BIESECKER, GREGORY  
; APPLICANT: GOLD, LARRY  
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF  
; NUMBER OF SEQUENCES: 157  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson & Bratschun, L.L.C.  
; STREET: 8400 E. Prentice Place #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: US  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,228B  
; FILING DATE: 12-FEBRUARY-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01739  
; FILING DATE: 30 JAN 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/595,335  
; FILING DATE: 1 FEB 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson, Esq.  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX50/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 60 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: RNA  
; FEATURE:  
; OTHER INFORMATION: All pyrimidines are 2'-F modified

US-09-023-228B-91

```

Query Match          72.4%;   Score 15.2;   DB 3;   Length 61;
Best Local Similarity 85.0%;   Pred. No. 26;
Matches 17;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

QY      2   tgtcggcggctggaatatc 21
          ||| ||| ||| ||| ||| |||
Db      47   TCACGAGCGGTGGAAGATC 28

RESULT      3
US-09-060-756-666
; Sequence 666, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:

```

ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4107.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2424 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-821-119-16

Query Match 72.4%; Score 15.2; DB 1; Length 2424;  
Best Local Similarity 85.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctcggcggtggaataatc 21  
||||| ||||| |||||

DB 463 TGTCGGGGGTGGACATATC 444

RESULT 5

US-08-617-860B-3/c  
; Sequence 3, Application US/08617860B  
; Patent No. 6133506  
; GENERAL INFORMATION:  
; APPLICANT: Tvpfer, R., Bautor, J., Bothmann, H., Filsak, E.,  
; APPLICANT: Hricke-Grandpierre, C., Klein, B., Martini, N.,  
; APPLICANT: Mlier, A., Schulte, W., Voetz, M., Walek, J.,  
; APPLICANT: Schell, J.  
; TITLE OF INVENTION: Promoters  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,860B  
; FILING DATE: 01-MAR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/02950  
; FILING DATE: 05-SEP-1994  
; APPLICATION NUMBER: DE P4329951.2  
; FILING DATE: 04-SEP-1993  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3350 base pairs  
; TYPE: Nucleic acid  
; STRANDEDNESS: Double stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Brassica napus  
; IMMEDIATE SOURCE:  
; LIBRARY: genomic Lambda FIX II  
; CLONE: BnAccaseg10  
; FEATURE:  
; NAME/KEY: Startcodon  
; LOCATION: 2611..2613  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(2611..2908, 3001..3341)  
US-08-617-860B-3

Query Match 72.4%; Score 15.2; DB 3; Length 3350;  
Best Local Similarity 85.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atctcggcggtggaataat 20  
||| ||| ||||| |||||

DB 600 ATGGCGGACGGTGGTAATAT 581

RESULT 6

US-07-721-761A-32  
; Sequence 32, Application US/07721761A  
; Patent No. 5475099  
; GENERAL INFORMATION:  
; APPLICANT: Vic. C. Knauf  
; APPLICANT: Gregory A. Thompson  
; TITLE OF INVENTION: Plant Fatty Acid Synthases  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 95616  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.7  
; SOFTWARE: Microsoft Word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/721,761A  
; FILING DATE: 19910626  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/568,493  
; FILING DATE: 15-AUGUST-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elizabeth Lassen  
; REGISTRATION NUMBER: 31,845  
; NAME: Donna E. Scherer  
; REGISTRATION NUMBER: 34,719  
; REFERENCE/DOCKET NUMBER: CGNE 76-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (916) 753-6313  
; TELEFAX: (916) 753-1510  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1533 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
US-07-721-761A-32

Query Match 71.4%; Score 15; DB 1; Length 1533;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcggcggtggaataat 17

DB 1117 GTCCGGCGGTGGAAA 1131

RESULT 7

US-07-978-687-32  
; Sequence 32, Application US/07978687  
; Patent No. 5510255  
; GENERAL INFORMATION:  
; APPLICANT: Vic. C. Knauf  
; APPLICANT: Gregory A. Thompson  
; TITLE OF INVENTION: Plant Fatty Acid Synthases  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: CA  
; COUNTRY: USA

ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/978,687  
FILING DATE: FEBRUARY 1, 1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05801  
FILING DATE: 15-AUGUST-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/558,493  
FILING DATE: 15-AUGUST-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/721,761  
FILING DATE: 26-JUNE-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1533 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
US-07-978-687-32

Query Match 71.4%; Score 15; DB 1; Length 1533;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcggcggtggaaa 17  
|||||  
Db 1117 GTCGGCGGTGGAAA 1131

RESULT 8  
US-08-471-791-12  
Sequence 12, Application US/08471791  
Patent No. 5723595  
GENERAL INFORMATION:  
APPLICANT: Thompson, Gregory A  
APPLICANT: Knauf, Vic C  
TITLE OF INVENTION: Plant Desaturases-Compositions and Uses  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: California  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.1  
SOFTWARE: Microsoft Word 5.1 (a)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,791  
FILING DATE: 6-JUNE-95  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/762,762  
FILING DATE: 16-SEPT-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01746  
FILING DATE: 14-MAR-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/615,784  
FILING DATE: 14-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/567,373  
FILING DATE: 13-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/494,106  
FILING DATE: 16-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Lassen, Elizabeth  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
NAME: Carl J. Schwedler  
REGISTRATION NUMBER: 36,924  
REFERENCE/DOCKET NUMBER: CGNE 69-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1533 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
US-08-471-791-12

Query Match 71.4%; Score 15; DB 1; Length 1533;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcggcggtggaaa 17  
|||||  
Db 1117 GTCGGCGGTGGAAA 1131

RESULT 9  
PCT-US91-01746-12  
Sequence 12, Application PC/TUS9101746  
GENERAL INFORMATION:  
APPLICANT: Thompson, Gregory A  
APPLICANT: Knauf, Vic C  
TITLE OF INVENTION: Plant Desaturases-Compositions and Uses  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: California  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01746  
FILING DATE: 19910314  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:









**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	16.8	80.0	288	9	BB607934	BB607934	BB607934
C	16.8	80.0	428	9	AV213384	AV213384	AV213384
	16.8	80.0	460	9	AW829820	AW829820	AW829820
C	16.8	80.0	613	12	BH532588	BH532588	BH532588
C	16.8	80.0	668	10	BG160433	BG160433	BG160433
C	16.8	80.0	792	10	BF627929	BF627929	BF627929
7	16.8	80.0	1278	12	B09467	B09467	B09467
8	16.8	80.0	1293	10	BM415118	BM415118	BM415118
C	16.4	78.1	1406	10	BE66361	BE66361	BE66361
10	16.2	77.1	161	12	BH205870	BH205870	BH205870
C	16.2	77.1	161	12	BH209293	BH209293	BH209293
12	16.2	77.1	161	12	BH209230	BH209230	BH209230
13	16.2	77.1	373	10	BG362171	BG362171	BG362171
14	16.2	77.1	423	10	BG409025	BG409025	BG409025
15	16.2	77.1	463	9	AW473489	AW473489	AW473489
16	16.2	77.1	506	10	BF082179	BF082179	BF082179
17	16.2	77.1	555	10	B1880600	B1880600	B1880600

COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenro-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki,  
Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh, M., Kitsumi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1. 428  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2410127F20"  
/clone\_lib="RIKEN full-length enriched, ES cells"  
/cell\_type="ES cells"  
/lab\_host="SOLR"  
/note="Site.1: XhoI; Site.2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 5.0 and subtraction to Rot = 25.0. Second strand  
cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGAGATTCTCGAGTTTCTTTTCTTTTNN 3']".

BASE COUNT 94 a 145 c 90 g 99 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 428;  
Best Local Similarity 90.0%; Pred. No. 8.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaaatc 20  
||||| ||||| |||||  
Db 86 ATGTCGGGGGTGGGAATAT 67

RESULT 3

AW829820 460 bp mRNA linear EST 10-MAY-2001  
LOCUS ra44g10.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita  
DEFINITION cDNA 5', mRNA sequence.  
ACCESSION AW829820  
VERSION AW829820.1 GI:7923632  
KEYWORDS EST.  
SOURCE southern root-knot nematode.  
ORGANISM Meloidogyne incognita  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
REFERENCE 1 (bases 1 to 460)

Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1. 288  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="E230008G20"  
/clone\_lib="RIKEN full-length enriched, 2 days pregnant  
adult female oviduct"  
/sex="female"  
/tissue\_type="oviduct"  
/dev\_stage="2 days pregnant adult"  
/lab\_host="DH10B"  
/note="Site.1: SalI; Site.2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGATTCTCGAGTTTCTTTTCTTTTNN 3']". cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FIC I."

BASE COUNT 53 a 67 c 111 g 57 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 288;  
Best Local Similarity 90.0%; Pred. No. 8.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgtcggcggtggaaatc 21  
||||| ||||| |||||  
Db 31 TGTCTGGGGGTGGGAATGTC 50

RESULT 2  
AV213384/c 428 bp mRNA linear EST 30-OCT-1999  
LOCUS AV213384 RIKEN full-length enriched, ES cells Mus musculus cDNA  
DEFINITION clone 2410127F20 3' similar to AF050165 Mus musculus protein  
phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION AV213384  
VERSION AV213384.1 GI:6152886  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 428)  
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Konno, H., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,  
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Owa, C., Ozawa, Y., Saio, H., Sano, M., Sato, K., Shibata, K., Shibata,  
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,  
Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y.,  
Watahiki, A., Watanabe, S., Yamamura, T., Yoshinaka, A., Yokota, T.,  
Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al. 1999)  
Unpublished (1999)

TITLE  
JOURNAL

## AUTHORS

McCarter, J., Clifton, S., Chlapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, W., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

The Washington Univ. Nematode EST Project, 1999  
Unpublished (1999)

## JOURNAL

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Uma Rao and David Bird

(david.bird@ncsu.edu) at North Carolina State University. DNA

Sequencing by: Washington University Genome Sequencing Center St.

Louis.

Seq primer: -40RP from Gibco

High quality sequence stop: 438.

Location/Qualifiers

## FEATURES

source

1..460

/organism="Meloiodogyne incognita"

/db\_xref="taxon:6306"

/clone\_lib="Bird-Rao Meloiodogyne incognita J2"

/dev\_stage="enriched for 2nd stage juveniles"

/lab\_host="XLORL"

/note="Vector: ZAP express - pBKCWV (Stratagene); Site\_1:

EcoRI; Site\_2: XhoI; Oligo (dT) primed library. cDNA was

constructed and cloned unidirectionally into the vector

within the 5' EcoRI and 3' XhoI sites. This library was

constructed by Dr. Uma Rao and Dr. David Bird at North

Carolina state University."

160 a 80 c 86 g 134 t

## BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 80.0%; Score 16.8; DB 9; Length 460;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcggcggtggaatatc 21

||||| ||| ||||| |||||

Db 198 TGTGAGGACGTGGAATATC 217

## RESULT 4

BH532588/c

LOCUS

BOGH87R BOGH Brassica oleracea genomic clone BOGH87, DNA

sequence. 613 bp DNA linear GSS 14-DEC-2001

ACCESION

BH532588

VERSION

BH532588.1 GI:17754954

KEYWORDS

GSS.

SOURCE

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 613)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOGH87TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

## Seq primer: TR

Class: sheared ends.

## FEATURES

source

Location/Qualifiers

1..613

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOGH87"

/clone\_lib="BOGH"

/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

176 a 113 c 106 g 218 t

## BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 80.0%; Score 16.8; DB 12; Length 613;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20

||||| ||| ||||| |||||

Db 266 ATGTCGGTGATGGAATAT 247

## RESULT 5

BG160433/c

LOCUS

DEFINITION

668 bp mRNA linear EST 06-FEB-2001

IMAGE:3580245 5' similar to SW:CRS3\_HORSE\_O19010 CYSTEINE-RICH

SECRETORY PROTEIN-3 PRECURSOR. [1]; mRNA sequence.

ACCESION

BG160433

VERSION

BG160433.1 GI:12694352

KEYWORDS

EST.

SOURCE

ORGANISM

Xenopus laevis

African clawed frog.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 668)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,

Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person

, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R. and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by N. Garrett, K. Ryan and A.M. Zorn,

(Wellcome/CRC Institute). DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 513.

Location/Qualifiers

1..668

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="IMAGE:3580245"

/clone\_lib="Wellcome CRC PRN3 head"

/tissue\_type="head, stage 30"

/lab\_host="DH10B (phage-resistant)"

/note="Vector: pBSRN3; Site\_1: NotI; Site\_2: EcoRI; cDNAs

were oligo-dT primed and directionally cloned. Staging

according to Nieuwkoop and Faber. Library was constructed

by N. Garrett, K. Ryan and A.M. Zorn, (Wellcome/CRC

Institute)."

197 a 162 c 143 g 165 t 1 others

## BASE COUNT

## ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 668;  
 Best Local Similarity 90.0%; Pred. No. 9.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgctcgccggtggaataatc 21  
 ||||| |||||||  
 Db 151 TGCTGACGGTGGAAATATC 132

## RESULT 6

BF627929 792 bp mRNA linear EST 22-OCT-2001  
 LOCUS HVSMEB0006K19f Hordeum vulgare seedling shoot EST library  
 DEFINITION HVSMEB0006K19f (Dehydration stress) Hordeum vulgare cDNA clone  
 HVSMEB0006K19f, mRNA sequence.

ACCESSION BF627929  
 VERSION BF627929.2 GI:13090236

## KEYWORDS

EST.

## SOURCE

barley.

## ORGANISM

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Hordeum.

1 (bases 1 to 792)

Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu

, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton

, R.D., Oates, R. and Main, D.

Development of a genetically and physically anchored EST resource

for barley genomics: Morex drought-stressed seedling shoot cDNA

library

Unpublished (2001)

On Dec 19, 2000 this sequence version replaced gi:11892087.

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hq bases = 392

Seq primer: AATTAACTCTCACTAAAGGG

High quality sequence start: 14

High quality sequence stop: 676.

Location/Qualifiers

1..792

/organism="Hordeum vulgare"

/cultivar="Morex"

/db\_xref="taxon:4513"

/clone="HVSMEB0006K19f"

/clone\_lib="Hordeum vulgare seedling shoot EST library

HVCNDA0002 (Dehydration stress)"

/tissue\_type="Seedling shoot"

/lab\_host="TJC121"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI;

Seeds were surface sterilized then germinated under axenic

conditions in the dark at room temperature on filter paper

with water, nystatin and cefotaxime in covered

crystallization dishes. Five-day old seedlings were

incubated at 90% RH for 24 hr. Shoots were then harvested,

total RNA was prepared, poly(A) RNA was purified, one

primary unamplified cDNA library was made, 600000 pfu were

in vivo excised to give pBluescript SK(-) cDNA phagemids.

These steps were performed in the TJ Close laboratory at

the University of California, Riverside (Choi, Close,

Fenton). Phagemids were plated and picked at the Clemson

University Genomics Institute (CUGI) (Begum, Palmer, Frisch,

Atkins and Wing). Plasmid DNA preparations, DNA

sequencing and sequence analysis were performed at CUGI

(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).

The sequence has been trimmed to remove vector sequence

and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders

Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)

Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.

(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

166 a 201 c 291 g 134 t

BASE COUNT 166 a 201 c 291 g 134 t

## ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 792;

Best Local Similarity 90.0%; Pred. No. 9.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgtcgccggtggaataatc 20

||||| |||||||

Db 556 AAGTGGCGGTGGAATAT 575

## RESULT 7

B09467

LOCUS

DEFINITION

B09467

VERSION

B09467.1

GI:2090587

KEYWORDS

GSS:

SOURCE

Arabidopsis thaliana

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 1278)

Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and

Ecker, J.

BAC End Sequences at ATGC

Unpublished (1997)

Other\_GSSs: F20P3-Sp6

Contact: Ecker J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA

19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenome.bio.upenn.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 38

High quality sequence stop: 702.

Location/Qualifiers

1..1278

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="F20P3"

/clone\_lib="IGF"

/sex="hermaphrodite"

/note="Vector: BelOACII; Site\_1: EcoRI; Site\_2: EcoRI;

Produced by Thomas Altmann"

349 a 254 c 268 g 405 t

BASE COUNT 349 a 254 c 268 g 405 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 12; Length 1278;

Best Local Similarity 90.0%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgctcgccggtggaataatc 21

||||| |||||||



Db 1094 TGTGGGGCGGTGATATATC 1113

# RESULT 8

BM415118 1293 bp mRNA linear EST 28-JAN-2002  
LOCUS OP20190 Mixed Stage EST's from Globodera pallida, the potato cyst  
DEFINITION nematode Globodera pallida cDNA, mRNA sequence.

BM415118

# ACCESSION

BM415118

# VERSION

BM415118.1 GI:18381473

# KEYWORDS

EST.

# SOURCE

ORGANISM Globodera pallida.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.  
1 (bases 1 to 1293)  
AUTHORS Heer, J., Sosinski, B., Pokrzywa, R.M., Warry, A. and Opperman, C.  
TITLE Mixed Stage EST's from Globodera pallida, the potato cyst nematode  
JOURNAL Unpublished (2001)  
COMMENT Contact: Opperman, C  
Center for the Biology of Nematode Parasitism  
NC State University; IACR-Rothamsted  
Campus Box 7616; Raleigh, NC 27695, USA  
Tel: 919.515.6699  
Fax: 919.515.9500  
Email: warthog@unity.ncsu.edu  
GT11-11PCN\_R\_H11\_GT11-11\_R\_092.ab1.

# FEATURES

source

1..1293  
/organism="Globodera pallida"  
/db\_xref="taxon:36090"  
/clone\_lib="Mixed Stage EST's from Globodera pallida, the  
potato cyst nematode"  
/note="Vector: lambda GT11; This is a collaborative effort  
between IACR-Rothamsted and North Carolina State  
University. The library was constructed from mixed stage  
G. pallida in lambda GT11 by Paul Burroughs,  
IACR-Rothamsted."

# BASE COUNT

ORIGIN

327 a 340 c 282 g 320 t 24 others  
Query Match 80.0%; Score 16.8; DB 10; Length 1293;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 atgtcggcggtggaatat 20  
||||||| |  
Db 115 ATGTCGGCGCTAGAAATAT 134

# RESULT 9

BI666361/C

# LOCUS

DEFINITION 603290686F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5325093 5',  
mRNA sequence.

# ACCESSION

BI666361

# VERSION

BI666361.1 GI:15580594

# KEYWORDS

EST.

# SOURCE

ORGANISM Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1406)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

# TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapps@mail.nih.gov  
Tissue procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

# CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

plate: LLAM1824 row: 1 column: 22

High quality sequence start: 18

High quality sequence stop: 75.

# FEATURES

source

1..1406  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5325093"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Salt;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 393 a 338 c 296 g 379 t

ORIGIN

Query Match 78.1%; Score 16.4; DB 10; Length 1406;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 gtctggcggtggaatat 20  
||||||| |  
Db 1184 GACGGCGGTGGAATAT 1167

# RESULT 10

BH205870

# LOCUS

DEFINITION BH205870 161 bp DNA linear GSS 24-OCT-2001  
Sml-54D8.TR Sml Schistosoma mansoni genomic clone Sml-54D8, DNA  
sequence.

# ACCESSION

BH205870

# VERSION

BH205870.1 GI:16381631

# KEYWORDS

GSS.

# SOURCE

ORGANISM

Schistosoma mansoni.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

# REFERENCE

1 (bases 1 to 161)

# AUTHORS

Shetty, J., Simpson, A., Malek, J., Koo, H., LoVerde, P.T. and El-Sayed  
, N.M.

# TITLE

Use of end sequences from Schistosoma mansoni (Puerto Rico strain)  
Sml BAC library for gene discovery and map construction

# JOURNAL

Unpublished (2001)

# COMMENT

Other\_GSSs: Sml-54D8.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
lo.edu).

Seq primer: M13 Rev

Class: BAC ends.

# FEATURES

source

1..161  
/organism="Schistosoma mansoni"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sml-54D8"  
/clone\_lib="Sml"  
/note="Vector: pBelOAC11; Site\_1: Hin dIII; Constructed  
in the laboratory of Dr. Denis Le Paslier at the Fondation  
Jean Bausset, CEPH, Paris, France. Briefly, Schistosoma

mansonii agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 32 c 41 g 44 t

BASE COUNT  
ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;  
Best Local Similarity 85.7%; Pred. No. 1.4e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21  
||||| ||||| ||||| ||

Db 16 ATGTCGGTCGGTTGAAATGTC 36

RESULT 11  
BH209290/c

LOCUS Sml-56P1.TF Sml Schistosoma mansoni genomic clone Sml-56P1, DNA  
DEFINITION Sml-56P1.TF Sml Schistosoma mansoni genomic clone Sml-56P1, DNA  
sequence.

ACCESSION BH209290

VERSION BH209290.1 GI:16387816

KEYWORDS GSS.

SOURCE Schistosoma mansoni.

ORGANISM Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 161)

REFERENCE Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed

AUTHORS 'N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)

Sml BAC library for gene discovery and map construction

Unpublished (2001)

Other\_GSSs: Sml-56P1.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Seq primer: M13 For

Class: BAC ends.

Location/Qualifiers

1..161

/organism="Schistosoma mansoni"

/strain="Puerto Rico"

/db\_xref="taxon:6183"

/clone="Sml-56P1"

/clone\_lib="Sml"

/note="Vector: pBelBAC11; Site\_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 41 c 32 g 44 t

BASE COUNT  
ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;

Best Local Similarity 85.7%; Pred. No. 1.4e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21

||||| ||||| ||||| ||

Db 146 ATGTCGGTCGGTTGAAATGTC 126

RESULT 12

BH209293

LOCUS Sml-56P1.TF Sml Schistosoma mansoni genomic clone Sml-56P1, DNA

DEFINITION Sml-56P1.TF Sml Schistosoma mansoni genomic clone Sml-56P1, DNA

sequence.

ACCESSION BH209293

VERSION BH209293.1 GI:16387822

KEYWORDS GSS.

SOURCE Schistosoma mansoni.

ORGANISM Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 161)

REFERENCE Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed

AUTHORS 'N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)

Sml BAC library for gene discovery and map construction

Unpublished (2001)

Other\_GSSs: Sml-56P1.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Seq primer: M13 Rev

Class: BAC ends.

Location/Qualifiers

1..161

/organism="Schistosoma mansoni"

/strain="Puerto Rico"

/db\_xref="taxon:6183"

/clone="Sml-56P1"

/clone\_lib="Sml"

/note="Vector: pBelBAC11; Site\_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 32 c 41 g 44 t

BASE COUNT

ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;

Best Local Similarity 85.7%; Pred. No. 1.4e+03;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21

||||| ||||| ||||| ||

Db 16 ATGTCGGTCGGTTGAAATGTC 36

RESULT 13

BG362171

LOCUS BG362171

DEFINITION BG362171

9B52f07.y1 Moss EST library PPG Physcomitrella patens cDNA clone

PEP\_SOURCE\_ID: 5', mRNA sequence.

373 bp mRNA linear EST 08-MAR-2001

```

ACCESSION   BG362171
VERSION     BG362171.1  GI:13251268
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1 (bases 1 to 373)
AUTHORS     Quatrano, R., Bashlades, S., Cove, D., Cumling, A., Knight, C., Clifton
            , S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood
            , K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
            Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.
            , Waterston, R. and Wilson, R.
TITLE       Leeds/Wash U Moss EST Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel.: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashlades as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 338.

FEATURES    1..373
            Location/Qualifiers
            /organism="Physcomitrella patens"
            /db_xref="taxon:3218"
            /clone="PEP_SOURCE ID:"
            /clone_lib="Moss EST library PPG"
            /tissue_type="gametophore: 30 day old tissue,
            ammonium-grown"
            /lab_host="DH10B"
            /note="Vector: pAMP1; Construction of the cDNA library was
            performed by Dr. W. Gregg Clark using a modification of
            the cDNA synthesis protocol developed in the laboratory of
            Dr. Michael Lovett by Dr. Yulia Korshunova (personal
            communication). First polyA + RNA was isolated from total
            gametophore RNA using oligo dT magnetic beads. Following
            this, first strand cDNA synthesis was performed on the
            bead-bound polyA + RNA, during which an oligonucleotide
            anchor sequence was incorporated onto the 5'-ends of the
            cDNA. PCR amplification was then used to synthesize the
            second strand, to amplify the double stranded DNA, and to
            incorporate dUTP containing sequences into the ends of the
            double stranded cDNA. This DNA was size selected and
            cloned into pAMP1 using the CloneAMP pAMP1 System (Life
            Technologies, GibcoBRL) for cloning amplification products
            by a non-restriction site dependant process. The cloning
            was directional based on sequence asymmetry introduced at
            the ends during PCR amplification. The 3' cDNA ends are
            proximal to the NotI site of the multiple cloning site in
            pAMP1. This annealing mixture was transformed into
            chemically competent DH10B cells and selected for
            ampicillin resistant growth. The resulting clones (about
            330,000) were pooled to make the library."
BASE COUNT  100 a  81 c  86 g  106 t
ORIGIN

Query Match      77.1%; Score 16.2; DB 10; Length 373;
Best Local Similarity 85.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 atgtcggcggtggaatac 21
    ||||| ||| ||||| |||
DB  21 ATGCTGCTGGCTGTGGAACATC 41

```

## RESULT 14

BG409025

LOCUS

DEFINITION

ACCESSION

VERSION

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..423

/organism="Physcomitrella patens"

/db\_xref="taxon:3218"

/clone="PEP\_SOURCE ID:PPG.CopyA-111218"

/clone\_lib="Moss EST library PPG"

/tissue\_type="gametophore: 30 day old tissue,

ammonium-grown"

/lab\_host="DH10B"

/note="Vector: pAMP1; Construction of the cDNA library was

performed by Dr. W. Gregg Clark using a modification of

the cDNA synthesis protocol developed in the laboratory of

Dr. Michael Lovett by Dr. Yulia Korshunova (personal

communication). First polyA + RNA was isolated from total

gametophore RNA using oligo dT magnetic beads. Following

this, first strand cDNA synthesis was performed on the

bead-bound polyA + RNA, during which an oligonucleotide

anchor sequence was incorporated onto the 5'-ends of the

cDNA. PCR amplification was then used to synthesize the

second strand, to amplify the double stranded DNA, and to

incorporate dUTP containing sequences into the ends of the

double stranded cDNA. This DNA was size selected and

cloned into pAMP1 using the CloneAMP pAMP1 System (Life

Technologies, GibcoBRL) for cloning amplification products

by a non-restriction site dependant process. The cloning

was directional based on sequence asymmetry introduced at

the ends during PCR amplification. The 3' cDNA ends are

proximal to the NotI site of the multiple cloning site in

pAMP1. This annealing mixture was transformed into

chemically competent DH10B cells and selected for

ampicillin resistant growth. The resulting clones (about

330,000) were pooled to make the library."

BASE COUNT 107 a 104 c 90 g 121 t

ORIGIN

Query Match

77.1%; Score 16.2; DB 10; Length 423;

Best Local Similarity

85.7%; Pred. No. 1.6e+03;

Matches 18; Conservative

0; Mismatches 3; Indels

0; Gaps

0;

QY 1 atgtcggcggtggaatac 21

||||| ||| ||||| |||

DB 21 ATGCTGCTGGCTGTGGAACATC 41

Best Local Similarity 85.7%; Pred. No. 1.7e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataac 21  
||||| ||| ||||| |||  
Db 86 ATGTCGGCTGTGGAACATC 106

RESULT 15  
AW473489  
LOCUS  
DEFINITION  
AW473489 463 bp mRNA linear EST 24-FEB-2000  
ha75f01.x1 NCI-CGAP Gas4 Homo sapiens cDNA clone IMAGE:2879545 3'  
similar to contains Alu repetitive element; contains element OFR  
repetitive element ;, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AW473489  
AW473489.1 GI:7043595  
EST.  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 463)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncilogap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 401.

FEATURES  
source  
1..463  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2879545"  
/clone\_lib="NCI-CGAP\_Gas4"  
/tissue\_type="poorly differentiated adenocarcinoma with  
signed ring cell features"  
/lab\_host="DH10B"  
/note="Organ: stomach; Vector: pCMV-SPORT6; Site:1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.69 kb. Life Technologies catalog #:  
11549-011"

BASE COUNT 119 a 121 c 91 g 132 t  
ORIGIN

Query Match 77.1%; Score 16.2; DB 9; Length 463;  
Best Local Similarity 85.7%; Pred. NO. 1.7e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataac 21  
||||| ||| ||||| |||  
Db 196 ATGTCGGGTGGGGGAATATC 216

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:38:07 ; Search time 3312.9 Seconds  
(without alignments)  
145.284 Million cell updates/sec

Title: US-09-235-416-4  
Perfect score: 23  
Sequence: 1 gaattctgcttcgtgttttca 23

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length	ID Description
-----			

1	18.8	81.7	72241	2	AC069180	AC069180 Homo sapi	
2	18.8	81.7	107744	9	AC013525	AC013525 Homo sapi	
C	3	18.8	81.7	139563	9	HSJ666F24	AL109660 Human DNA
4	18.8	81.7	148184	2	AC017777	AC017777 Drosophil	
5	18.8	81.7	155047	2	AL390120	AL390120 Homo sapi	
C	6	18.8	81.7	177908	2	AC073977	AC073977 Homo sapi
7	18.8	81.7	195755	3	AC010009	AC010009 Drosophil	
8	18.8	81.7	212739	2	AC018630	AC018630 Homo sapi	
9	18.8	81.7	220761	2	AL670227	Mus muscu	
C	10	18.8	81.7	295566	3	AE003597	Drosophil
11	18.4	80.0	1305	6	AX122257	Sequence	
C	12	18.4	80.0	112618	9	AL390862	Human DNA
C	13	18.4	80.0	171943	9	AC012055	Homo sapi
C	14	18.4	80.0	179667	2	AC106895	Homo sapi
C	15	18.4	80.0	349980	6	AX127149	Sequence
16	18.2	79.1	1350	9	HSRNRAGA		
17	18.2	79.1	1582	9	BC006433	Homo sapi	
18	18.2	79.1	1607	9	BC009990	Homo sapi	
19	18.2	79.1	1610	9	HS041654	Human adeno	
20	18.2	79.1	2418	5	CS023186	Crotalus sc	
21	18.2	79.1	2620	5	VAMM12		
22	18.2	79.1	2769	5	U01027		
23	18.2	79.1	3056	5	VAMMOL		
24	18.2	79.1	3224	5	U01026		
25	18.2	79.1	3571	5	TRUGTGPLAB		
C	26	18.2	79.1	6435	14	PVXX3	
C	27	18.2	79.1	22204	1	AE008842	Salmonell
28	18.2	79.1	57714	2	AC107999	Homo sapi	
29	18.2	79.1	70374	2	AC100649	Mus muscu	
30	18.2	79.1	89177	9	AL356000	Human DNA	
31	18.2	79.1	110000	2	AL390072_2	Continuation (3 of	
C	32	18.2	79.1	160626	9	AC006374	Homo sapi
33	18.2	79.1	160705	9	HS61A9	Human DNA	
34	18.2	79.1	185848	2	AC025929	Homo sapi	
35	18.2	79.1	189371	2	AC023125	Homo sapi	
36	18.2	79.1	191748	2	AL662902	Mus muscu	
C	37	18.2	79.1	191772	2	AC022825	Homo sapi
38	18.2	79.1	193581	2	AL670941	Mus muscu	
C	39	18.2	79.1	200475	9	AC079271	Homo sapi
40	18.2	79.1	200618	9	AC009061	Homo sapi	
C	41	18.2	79.1	200724	9	AL139382	Human DNA
C	42	18.2	79.1	209262	2	AC021091	Homo sapi
C	43	18.2	79.1	210385	2	AC009627	Homo sapi
C	44	18.2	79.1	215399	9	AL160278	Human DNA
45	18.2	79.1	225401	2	AL671011	Mus muscu	

ALIGNMENTS

RESULT	1	AC069180	72241 bp	DNA	linear	HTG 21-MAY-2000
AC069180	LOCUS	AC069180	Homo sapiens chromosome 1 clone -2026P8 map 1, LOW-PASS SEQUENCE			
DEFINITION		AC069180	Homo sapiens chromosome 1, clone -2026P8			
ACCESSION		AC069180	HTG; HTGS_PHASE0.			
VERSION		AC069180.1	GI:7960385			
KEYWORDS			human.			
SOURCE			human.			
ORGANISM			Homo sapiens			
REFERENCE			Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE			1 (bases 1 to 72241)			
JOURNAL			Birren, B., Linton, L., Nusbaum, C. and Lander, E.			
REFERENCE			2 (bases 1 to 72241)			
AUTHORS			Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,			
			Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,			
			Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,			
			Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,			
			Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,			
			Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,			

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Olivier, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (21-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L6854  
 Center clone name: 2026\_P\_8  
 -----

\* NOTE: This record contains 93 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 658: contig of 658 bp in length  
 \* 659 758: gap of 100 bp  
 \* 759 1448: contig of 690 bp in length  
 \* 1449 1548: gap of 100 bp  
 \* 1549 2245: contig of 697 bp in length  
 \* 2246 2345: gap of 100 bp  
 \* 2346 3039: contig of 694 bp in length  
 \* 3040 3139: gap of 100 bp  
 \* 3140 3798: contig of 659 bp in length  
 \* 3799 3898: gap of 100 bp  
 \* 3899 4555: contig of 657 bp in length  
 \* 4556 4655: gap of 100 bp  
 \* 4656 5337: contig of 682 bp in length  
 \* 5338 5437: gap of 100 bp  
 \* 5438 6106: contig of 669 bp in length  
 \* 6107 6206: gap of 100 bp  
 \* 6207 6872: contig of 666 bp in length  
 \* 6873 6972: gap of 100 bp  
 \* 6973 7651: contig of 679 bp in length  
 \* 7652 7751: gap of 100 bp  
 \* 7752 8429: contig of 678 bp in length  
 \* 8430 8529: gap of 100 bp  
 \* 8530 9195: contig of 666 bp in length  
 \* 9196 9295: gap of 100 bp  
 \* 9296 9965: contig of 670 bp in length  
 \* 9966 10065: gap of 100 bp  
 \* 10066 10738: contig of 673 bp in length  
 \* 10739 10838: gap of 100 bp  
 \* 10839 11534: contig of 696 bp in length  
 \* 11535 11634: gap of 100 bp  
 \* 11635 12340: contig of 706 bp in length  
 \* 12341 12440: gap of 100 bp  
 \* 12441 13111: contig of 671 bp in length  
 \* 13112 13211: gap of 100 bp  
 \* 13212 13871: contig of 660 bp in length  
 \* 13872 13971: gap of 100 bp  
 \* 13972 14641: contig of 670 bp in length  
 \* 14642 14741: gap of 100 bp  
 \* 14742 15422: contig of 681 bp in length  
 \* 15423 15522: gap of 100 bp  
 \* 15523 16199: contig of 677 bp in length  
 \* 16200 16299: gap of 100 bp  
 \* 16300 16973: contig of 674 bp in length  
 \* 16974 17073: gap of 100 bp  
 \* 17074 17754: contig of 681 bp in length  
 \* 17755 17854: gap of 100 bp  
 \* 17855 18540: contig of 686 bp in length  
 \* 18541 18640: gap of 100 bp  
 \* 18641 19293: contig of 653 bp in length  
 \* 19294 19393: gap of 100 bp  
 \* 19394 20094: contig of 701 bp in length  
 \* 20095 20194: gap of 100 bp  
 \* 20195 20880: contig of 686 bp in length  
 \* 20881 20980: gap of 100 bp  
 \* 20981 21662: contig of 682 bp in length  
 \* 21663 21762: gap of 100 bp  
 \* 21763 22418: contig of 656 bp in length  
 \* 22419 22518: gap of 100 bp  
 \* 22519 23189: contig of 671 bp in length  
 \* 23190 23289: gap of 100 bp  
 \* 23290 23934: contig of 645 bp in length  
 \* 23935 24034: gap of 100 bp  
 \* 24035 24713: contig of 679 bp in length  
 \* 24714 24813: gap of 100 bp  
 \* 24814 25488: contig of 675 bp in length  
 \* 25489 25588: gap of 100 bp  
 \* 25589 26261: contig of 673 bp in length  
 \* 26262 26361: gap of 100 bp  
 \* 26362 27036: contig of 665 bp in length  
 \* 27037 27156: gap of 100 bp  
 \* 27157 27815: contig of 689 bp in length  
 \* 27816 27915: gap of 100 bp  
 \* 27916 28577: contig of 662 bp in length  
 \* 28578 28677: gap of 100 bp  
 \* 28678 29374: contig of 697 bp in length  
 \* 29375 29474: gap of 100 bp  
 \* 29475 30155: contig of 681 bp in length  
 \* 30156 30255: gap of 100 bp  
 \* 30256 30942: contig of 687 bp in length  
 \* 30943 31042: gap of 100 bp  
 \* 31043 31716: contig of 674 bp in length  
 \* 31717 31816: gap of 100 bp  
 \* 31817 32491: contig of 675 bp in length  
 \* 32492 32591: gap of 100 bp  
 \* 32592 33266: contig of 675 bp in length  
 \* 33267 33366: gap of 100 bp  
 \* 33367 34031: contig of 665 bp in length  
 \* 34032 34131: gap of 100 bp  
 \* 34132 34808: contig of 677 bp in length  
 \* 34809 34908: gap of 100 bp  
 \* 34909 35587: contig of 679 bp in length  
 \* 35588 35687: gap of 100 bp  
 \* 35688 36379: contig of 692 bp in length  
 \* 36380 36479: gap of 100 bp  
 \* 36480 37121: contig of 642 bp in length  
 \* 37122 37221: gap of 100 bp  
 \* 37222 37915: contig of 694 bp in length  
 \* 37916 38015: gap of 100 bp  
 \* 38016 38720: contig of 705 bp in length  
 \* 38721 38820: gap of 100 bp  
 \* 38821 39514: contig of 694 bp in length  
 \* 39515 39614: gap of 100 bp  
 \* 39615 40284: contig of 670 bp in length  
 \* 40285 40384: gap of 100 bp  
 \* 40385 41052: contig of 668 bp in length

TITLE  
 JOURNAL  
 COMMENT

\* 41053 41152: gap of 100 bp  
\* 41153 41829: contig of 677 bp in length  
\* 41830 41929: gap of 100 bp  
\* 41930 42621: contig of 692 bp in length  
\* 42622 42721: gap of 100 bp  
\* 42722 43401: contig of 680 bp in length  
\* 43402 43501: gap of 100 bp  
\* 43502 44177: contig of 676 bp in length  
\* 44178 44277: gap of 100 bp  
\* 44278 44955: contig of 678 bp in length  
\* 44956 45055: gap of 100 bp  
\* 45056 45750: contig of 695 bp in length  
\* 45751 45850: gap of 100 bp  
\* 45851 46530: contig of 680 bp in length  
\* 46531 47326: contig of 696 bp in length  
\* 47327 47426: gap of 100 bp  
\* 47427 48123: contig of 697 bp in length  
\* 48124 48223: gap of 100 bp  
\* 48224 48909: contig of 686 bp in length  
\* 48910 49009: gap of 100 bp  
\* 49010 49680: contig of 671 bp in length  
\* 49681 49780: gap of 100 bp  
\* 49781 50446: contig of 666 bp in length  
\* 50447 50546: gap of 100 bp  
\* 50547 51224: contig of 678 bp in length  
\* 51225 51324: gap of 100 bp  
\* 51325 52004: contig of 680 bp in length  
\* 52005 52104: gap of 100 bp  
\* 52105 52771: contig of 667 bp in length  
\* 52772 52871: gap of 100 bp  
\* 52872 53551: contig of 680 bp in length  
\* 53552 53651: gap of 100 bp

Query Match 81.7%; Score 18.8; DB 2; Length 72241;

Best Local Similarity 90.9%; Pred. No. 2.7e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 aattctgcttcgtctgtttcca 23

Db 36923 ACTTCTGCTTTCGTCTGTTTCA 36944

RESULT 2

AC013525 LOCUS 107744 bp DNA linear HTG 13-JUL-2000  
AC013525 DEFINITION Homo sapiens clone RP11-114H8, LOW-PASS SEQUENCE SAMPLING.  
AC013525 ACCESSION  
AC013525.3 GI:9113253  
HTG: HTGS\_PHASE0.  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 107744)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

REFERENCE

AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gaidyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6468112.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1925  
Center clone name: 114\_H\_8  
-----

\* NOTE: This record contains 111 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 915: contig of 915 bp in length  
\* 916 1015: gap of 100 bp  
\* 1016 1918: contig of 903 bp in length  
\* 1919 2018: gap of 100 bp  
\* 2019 2919: contig of 901 bp in length  
\* 2920 3019: gap of 100 bp  
\* 3020 3930: contig of 911 bp in length  
\* 3931 4030: gap of 100 bp  
\* 4031 4904: contig of 874 bp in length  
\* 4905 5004: gap of 100 bp  
\* 5005 5905: contig of 901 bp in length  
\* 5906 6005: gap of 100 bp  
\* 6006 6913: contig of 908 bp in length  
\* 6914 7013: gap of 100 bp  
\* 7014 7928: contig of 915 bp in length  
\* 7929 8028: gap of 100 bp  
\* 8029 8914: contig of 886 bp in length  
\* 8915 9014: gap of 100 bp  
\* 9015 9917: contig of 903 bp in length  
\* 9918 10017: gap of 100 bp  
\* 10018 10923: contig of 906 bp in length  
\* 10924 11023: gap of 100 bp  
\* 11024 11917: contig of 894 bp in length  
\* 11918 12017: gap of 100 bp  
\* 12018 12954: contig of 937 bp in length  
\* 12955 13054: gap of 100 bp  
\* 13055 13936: contig of 882 bp in length  
\* 13937 14036: gap of 100 bp  
\* 14037 14948: contig of 912 bp in length  
\* 14949 15048: gap of 100 bp  
\* 15049 15948: contig of 900 bp in length  
\* 15949 16048: gap of 100 bp  
\* 16049 16956: contig of 908 bp in length  
\* 16957 17056: gap of 100 bp  
\* 17057 17961: contig of 905 bp in length  
\* 17962 18061: gap of 100 bp  
\* 18062 18970: contig of 909 bp in length  
\* 18971 19070: gap of 100 bp  
\* 19071 19979: contig of 909 bp in length  
\* 19980 20079: gap of 100 bp  
\* 20080 20958: contig of 879 bp in length  
\* 20959 21058: gap of 100 bp  
\* 21059 21957: contig of 899 bp in length  
\* 21958 22057: gap of 100 bp  
\* 22058 22951: contig of 894 bp in length  
\* 22952 23051: gap of 100 bp

\* 23052 23940: contig of 889 bp in length  
 \* 23941 24040: gap of 100 bp  
 \* 24041 24941: contig of 901 bp in length  
 \* 24942 25041: gap of 100 bp  
 \* 25042 25965: contig of 924 bp in length  
 \* 25966 26065: gap of 100 bp  
 \* 26066 26956: contig of 891 bp in length  
 \* 26957 27056: gap of 100 bp  
 \* 27057 27966: contig of 910 bp in length  
 \* 27967 28066: gap of 100 bp  
 \* 28067 28973: contig of 907 bp in length  
 \* 28974 29073: gap of 100 bp  
 \* 29074 29988: contig of 915 bp in length  
 \* 29989 30088: gap of 100 bp  
 \* 30089 31011: contig of 923 bp in length  
 \* 31012 31111: gap of 100 bp  
 \* 31112 32020: contig of 909 bp in length  
 \* 32021 32120: gap of 100 bp  
 \* 32121 33026: contig of 906 bp in length  
 \* 33027 33126: gap of 100 bp  
 \* 33127 34032: contig of 906 bp in length  
 \* 34033 34132: gap of 100 bp  
 \* 34133 35038: contig of 906 bp in length  
 \* 35039 35138: gap of 100 bp  
 \* 35139 36058: contig of 920 bp in length  
 \* 36059 36158: gap of 100 bp  
 \* 36159 37057: contig of 899 bp in length  
 \* 37058 37157: gap of 100 bp  
 \* 37158 38058: contig of 901 bp in length  
 \* 38059 38158: gap of 100 bp  
 \* 38159 39076: contig of 918 bp in length  
 \* 39077 39176: gap of 100 bp  
 \* 39177 40089: contig of 913 bp in length  
 \* 40090 40189: gap of 100 bp  
 \* 40190 41105: contig of 916 bp in length  
 \* 41106 41205: gap of 100 bp  
 \* 41206 42107: contig of 902 bp in length  
 \* 42108 42207: gap of 100 bp  
 \* 42208 43095: contig of 888 bp in length  
 \* 43096 43195: gap of 100 bp  
 \* 43196 44090: contig of 895 bp in length  
 \* 44091 44190: gap of 100 bp  
 \* 44191 45083: contig of 893 bp in length  
 \* 45084 45183: gap of 100 bp  
 \* 45184 46091: contig of 908 bp in length  
 \* 46092 46191: gap of 100 bp  
 \* 46192 47091: contig of 900 bp in length  
 \* 47092 47191: gap of 100 bp  
 \* 47192 48109: contig of 918 bp in length  
 \* 48110 48209: gap of 100 bp  
 \* 48210 49095: contig of 886 bp in length  
 \* 49096 49195: gap of 100 bp  
 \* 49196 50097: contig of 902 bp in length  
 \* 50098 50197: gap of 100 bp  
 \* 50198 51097: contig of 900 bp in length  
 \* 51098 51197: gap of 100 bp  
 \* 51198 52137: contig of 940 bp in length  
 \* 52138 52237: gap of 100 bp  
 \* 52238 53142: contig of 905 bp in length  
 \* 53143 53242: gap of 100 bp  
 \* 53243 54135: contig of 893 bp in length  
 \* 54136 54235: gap of 100 bp  
 \* 54236 55135: contig of 900 bp in length  
 \* 55136 55235: gap of 100 bp  
 \* 55236 56107: contig of 872 bp in length  
 \* 56108 56207: gap of 100 bp  
 \* 56208 57100: contig of 893 bp in length  
 \* 57101 57200: gap of 100 bp  
 \* 57201 58109: contig of 909 bp in length  
 \* 58110 58209: gap of 100 bp  
 \* 58210 59113: contig of 904 bp in length  
 \* 59114 59213: gap of 100 bp  
 \* 59214 60122: contig of 909 bp in length

\* 60123 60222: gap of 100 bp  
 \* 60223 61095: contig of 873 bp in length  
 \* 61096 61195: gap of 100 bp  
 \* 61196 62104: contig of 909 bp in length  
 \* 62105 62204: gap of 100 bp  
 \* 62205 63123: contig of 919 bp in length  
 \* 63124 63223: gap of 100 bp  
 \* 63224 64141: contig of 918 bp in length  
 \* 64142 64241: gap of 100 bp  
 \* 64242 65137: contig of 896 bp in length  
 \* 65138 65237: gap of 100 bp  
 \* 65238 66154: contig of 917 bp in length  
 \* 66155 66254: gap of 100 bp  
 \* 66255 67090: contig of 836 bp in length  
 \* 67091 67190: gap of 100 bp  
 \* 67191 68081: contig of 891 bp in length  
 \* 68082 68181: gap of 100 bp  
 \* 68182 69095: contig of 914 bp in length  
 \* 69096 69195: gap of 100 bp  
 \* 69196 70073: contig of 878 bp in length  
 \* 70074 70173: gap of 100 bp  
 \* 70174 71071: contig of 898 bp in length  
 \* 71072 71171: gap of 100 bp  
 \* 71172 72065: contig of 894 bp in length  
 \* 72066 72165: gap of 100 bp  
 \* 72166 73083: contig of 918 bp in length

Query Match 81.7%; Score 18.8; DB 2; Length 107744;  
 Best Local Similarity 90.9%; Pred. NO. 2.6e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aattcgttcgtcgtgttttca 23  
 | ||||| ||||| ||||| |||||  
 Db 39895 ACTTCGCTTGTGCTTTTCA 39916

RESULT 3

HSJ566F24/c  
 LOCUS HSJ566F24 139563 bp DNA linear PRI 09-FEB-2002  
 DEFINITION Human DNA sequence from clone RP4-666F24 on chromosome lp13.1-13.3,  
 complete sequence.  
 ACCESSION AL109660  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 139563)  
 AUTHORS Frankland, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 COMMENT On Feb 11, 2002 this sequence version replaced gi:15594038.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence





```

source
misc_feature
1..18738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-521M12"
/clone_lib="RPC1-11.2"
1..18738
/feature="assembly_fragment:00869
fragment_chain:1
clone_end:Sp6
vector_side:left"
18839..22388
/feature="assembly_fragment:00738
fragment_chain:1"
22489..24572
/feature="assembly_fragment:01040
fragment_chain:1"
24673..28174
/feature="assembly_fragment:01064
fragment_chain:2"
28275..30525
/feature="assembly_fragment:01062
fragment_chain:2"
30626..50010
/feature="assembly_fragment:00964
fragment_chain:2"
50111..52235
/feature="assembly_fragment:00681
fragment_chain:2"
52356..58621
/feature="assembly_fragment:00865
fragment_chain:2"
58722..65489
/feature="assembly_fragment:00120
fragment_chain:3"
65590..73034
/feature="assembly_fragment:00668
fragment_chain:3"
73135..76041
/feature="assembly_fragment:00777
fragment_chain:4"
76142..87642
/feature="assembly_fragment:01128
fragment_chain:4"
87743..94188
/feature="assembly_fragment:00996
fragment_chain:5"

```

Sequencing vector: M13; 32%  
Chemistry: Dye-primer ET; 32% of reads  
Assembly: Dye-terminator Big Dye; 67% of reads  
Consensus quality: 172974 bases at least Q40  
Consensus quality: 174626 bases at least Q30  
Insert size: 160000; agarose-fp  
Insert size: 177308; sum-of-coverage  
Quality coverage: 12.66 in Q20 bases; agarose-fp  
Quality coverage: 11.47 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1146: contig of 1146 bp in length  
\* 1147 1246: gap of unknown length  
\* 1247 2420: contig of 1174 bp in length  
\* 2421 2520: gap of unknown length  
\* 2521 4176: contig of 1656 bp in length  
\* 4177 4276: gap of unknown length  
\* 4277 5991: contig of 1715 bp in length  
\* 5992 6091: gap of unknown length  
\* 6092 9107: contig of 3016 bp in length  
\* 9108 9207: gap of unknown length  
\* 9208 176414: contig of 167207 bp in length  
\* 176415 176514: gap of unknown length  
\* 176515 177908: contig of 1394 bp in length.

FEATURES

source  
1. 177908  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-253H16"  
1. 1146  
/note="assembly\_name:Contig21"  
1247. 2420  
/note="assembly\_name:Contig22"  
2521. 4176  
/note="assembly\_name:Contig28"  
4277. 5991  
/note="assembly\_name:Contig29"  
6092. 9107  
/note="assembly\_name:Contig33"  
9208. 176414  
/note="assembly\_name:Contig34"  
clone\_end:SP6  
vector\_side:left  
176515..177908  
/note="assembly\_name:Contig6"  
BASE COUNT 54431 a 35157 c 34279 g 53411 t 630 others  
ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 177908;  
Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattctgtctgcgtgtttc 22

||||||| ||| |||||

Db 65058 GAATTCCTCTTGTGCTTTC 65037

RESULT

AC010009

LOCUS

DEFINITION

195755 bp DNA linear INV 03-JAN-2002

Drosophila melanogaster 3L BAC RPC198-23D16 (Roswell Park Cancer

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS

TITLE

Institute Drosophila BAC library) complete sequence.

AC010009  
AC010009.5 GI:16418028  
HTG.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 195755)

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,

Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,

Ayele,M.A., Scott,G.S., Worley,K.W., Amaratunga,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhaý,C.,

Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenbush,K.,

Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,

Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,

Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,

Liu,W., Mattel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,

Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,

Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,

Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 195755)

Worley,K.C.

Direct Submission

Submitted (11-SEP-1999)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 195755)

Worley,K.C., Adams,C.,

Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brivava,M., Brown,E., Brown,N., Bryant,N.P., Buhaý,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,

Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W.,

Louisege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,

Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,

Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmami,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Watlington,S., Williams,G., Williams,A., Wleczek,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Direct Submission

## JOURNAL

Submitted (25-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

4 (bases 1 to 195755)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE  
JOURNAL

Submitted (27-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

5 (bases 1 to 195755)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE  
JOURNAL

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

6 (bases 1 to 195755)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Query Match 81.7%; Score 18.8; DB 3; Length 195755;  
 Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttc 22  
 ||||| ||||| ||||| |||||

Db 37603 GAATTCGCTTCTCTGTTTC 37624

## RESULT 8

AC018630 212739 bp DNA linear HTG 13-OCT-2001  
 LOCUS Homo sapiens chromosome 12 clone RP11-673D15, WORKING DRAFT  
 DEFINITION SEQUENCE, 2 unordered pieces.  
 AC018630  
 AC018630.33 GI:16041329  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS

Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 212739)  
 Muzny, D.M., Adams, C.C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J.,  
 Benton, J., Blum, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
 Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
 Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,  
 Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojuben, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshchik, N.,  
 Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Groves, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 212739)  
 Worley, K.C.  
 Direct Submission  
 Submitted (15-DEC-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 11, 2001 this sequence version replaced gi:14787084.

TITLE  
JOURNAL  
AUTHORS  
JOURNAL

## COMMENT

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HMG  
 Center clone name: HMG  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; M77789  
 Chemistry: Dye-terminator Big Dye; 7% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 219183 bases at least Q40  
 Consensus quality: 221416 bases at least Q30  
 Consensus quality: 222879 bases at least Q20  
 Estimated insert size: 215919; sum-of-contigs estimation  
 Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 151372: contig of 151372 bp in length  
 \* 151373 151472: gap of unknown length  
 \* 151473 212739: contig of 61267 bp in length.

FEATURES  
Source

Location/Qualifiers  
 1..212739  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-673D15"

BASE COUNT 70140 a 40132 c 36532 g 65833 t 102 others  
 ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 212739;  
 Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaattcgtgctgctgttttc 22

Db 96992 GAATTCGCTGCTCCATGTTTC 97013

## RESULT 9

AL670227 220761 bp DNA linear HTG 30-JAN-2002  
 LOCUS Mus musculus chromosome 4 clone RP23-317N1, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, in unordered pieces.

ACCESSION AL670227.2 GI:18477084

VERSION AL670227

KEYWORDS HTG: HTGS\_PHASE1.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Plumb, B.

Direct Submission

Submitted (23-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 1, 2002 this sequence version replaced gi:18307364.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

----- Project Information

Center project name: BM317N1

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 219697 bases at least Q40

Consensus quality: 219887 bases at least Q30

Consensus quality: 220081 bases at least Q20

Insert size: 220261; sum-of-contigs

Insert size: 213159; 11.3% error; agarose-fp

Quality coverage: 13.01x in Q20 bases; sum-of-contigs Quality

coverage: 13.51x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

FEATURES  
source

Location/Qualifiers  
 1..220761  
 /organism="Mus musculus"

```

/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-317N1"
/clone_1lb="RPC1-23"
1.129995
/note="assembly_fragment:03451
fragment_chain:1"
130096.133380
/note="assembly_fragment:02708
fragment_chain:1"
133481.150170
/note="assembly_fragment:00061
fragment_chain:1"
150271.156539
/note="assembly_fragment:01326
fragment_chain:1"
156640.208913
/note="assembly_fragment:02733
fragment_chain:1"
209014.220761
/note="assembly_fragment:05258
fragment_chain:1"
clone_end:SP6
vector_side:right"
BASE COUNT 57530 a 54793 c 52961 g 54975 t 502 others
ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 220761;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattcctgcctcgtctgtttc 22
|||||
Db 219218 GAATTCCTGCTCTCTGCTTTC 219239

RESULT 10
AE003597/c
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386036 section 7
of 9, complete sequence.
ACCESSION
AE003597 AE002647
VERSION
AE003597.1 GI:7296525
KEYWORDS
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 295566)
Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,J.H., Blaziel,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Fosiier,C., Gabriellian,A.E., Garg,N.S.,
Fleischmann,W., Fostier,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Pan,S.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE
JOURNAL
MEDLINE
REFERENCE
2 (bases 1 to 295566)
AUTHORS
Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
TITLE
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
JOURNAL
FEATURES
source
location/Qualifiers
1.295566
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
join(<27808..27909,28034..28318,28822..>29085)
/gene="CG12673"
/product="CT35415"
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
<27808..>29085
/gene="CG12673"
/map="79D4-79E1"
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
join(27808..27909,28034..28318,28822..>29085)
/gene="CG12673"
/note="CG12673 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
/protein_id="AAF51810.1"
/db_xref="GI:7296526"
/translation="MGREVEGRRAFGPRGAFPAQCIRVHVSNIDIAANDTMRLLYMY
HAQDPHGVSVRPGTLPDPAFAFRPRMVLQRAQLPMSPTHDERVRLVLRNEDVE
LPAGDPLFCMKFLEIDNRKHLIRYPIYDSSSVHYLQHTLHCOCGAHALEEE
MARQGRPCIGARSIPICLAICNAIVASWRSQSVISIRERONLPYRQCNSCPQT"
join(<35333..35500,35594..35756,35909..>35937)
/gene="CG7495"
/product="CT23023"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
<35333..>35937
/gene="CG7495"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
join(35333..35500,35594..35756,35909..>35937)
/gene="CG7495"
/note="CG7495 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
/protein_id="AAF51811.1"
/db_xref="GI:7296527"
/translation="MDPNWRHIIPPQKRVVSEQCIEDCTGYAFPOQGINIFAVMMR"

```

THQIGKVKLRQIRQTELPPIAHDSNDIVAYQDFRRLPQSVHSPGDRLLAECIYDS  
SSRAITIGEDVEFPALS  
Join(<48427. .48474,48682. .48989,49294. .>493118)  
/gene="CG14461"  
/product="CT34144"  
/db\_xref="FLYBASE:FBan0014461"  
/db\_xref="FLYBASE:FBgn0037155"  
/evidence=not\_experimental  
<48427. .>493118  
/gene="CG14461"  
/map="79E1-79E1"  
/db\_xref="FLYBASE:FBan0014461"  
/db\_xref="FLYBASE:FBgn0037155"  
/evidence=not\_experimental  
Join(48427. .48474,48682. .48989,49294. .493118)  
/gene="CG14461"  
/note="CG14461 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0014461"  
/db\_xref="FLYBASE:FBgn0037155"  
/evidence=not\_experimental  
/protein\_id="AAF51812.1"  
/db\_xref="GI:7296528"  
/translation="MWAEISDGTLCQLSGSETELEGYSINVTYKKHRRCKKPEPL  
APPTERTAPPASDLSLSEPLVHEDLNNIIEGAARSSSSADVHLSRSGSRHFISC  
LLWLGAASSWLLMLLRTPTPTTC"  
complement(join(64112. .66736,66987. .69059,69117. .69239))  
/gene="BcDNA:GH03694"  
/product="CT26020"  
/db\_xref="FLYBASE:FBan0009063"  
/db\_xref="FLYBASE:FBgn0028500"  
complement(<64112. .>69239)  
/gene="BcDNA:GH03694"  
/note="CG9063"  
/map="79E1-79E1"  
/db\_xref="FLYBASE:FBan0009063"  
/db\_xref="FLYBASE:FBgn0028500"  
/gene="BcDNA:GH03694"  
/note="BcDNA:GH03694 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009063"  
/db\_xref="FLYBASE:FBgn0028500"  
/protein\_id="AAF51813.1"  
/db\_xref="GI:7296529"  
/translation="MYEPVGPVKRVGLALPGESASIRHICCDVAKILVAAGDDFLGI  
WYANPLPIAFYRTEDSLQYGNOLIVKWPDSROLALLTAGSLLLYQLDFAENG  
GILQIDLPPLASLRKDSAELEFKENIPRLSRELGCSVTLSGVTTCISLSELLAT  
OSCELLRLQWTELEHAENDLELPALSSIKLRDIPFYVQQQOQSARNVPLNDRSYA  
SLEYSFFIGGCAAFSDRAAFLIANHURFETDHHGFWPVDVEDASVCSVNHKFRLL  
AYGQESSAVYVADDATGGLEFFSHRLITENILPDSLGSLNKLWSPDGCVLAYSWT  
NGLSLWSTFGALLSTLWDFGLNVLQCNPLKIRRLWSTEGYQLFMKLHPKED  
KSNVLQGFVKASLSMNPMTTSPHILLQGDCLYLQGNLLELTYSAGSHGTFPSGL  
GSDEDIQDGLCELEKQSPHTGSILTESKYVTVLQPLNYAATNWPRIYAADPDGLH  
LAVAGRTGLAHYSLVTRWKLGNESQEKDFVSGLLWHGVYMGCSLLDRDEL  
RCYPADCKLDNQYGHKLVQVAPVISLNSFRHQLIVITADGVLSLFNMSKNSAYLADIE  
CAYEDLTASICHPACIVSLVTNLKELKPGQLGQDAETIIVNVCGRILMIQORDA  
GEQVNTLTATCSLSEVFLWSLERSLCAMDCLWLYSGAHGMVWLPILPPGRERR  
EEQGLPDLFVKSRLMKRLMFLPLVLYFDNVIIVLGVNESTLVANEQVSHFSL  
PFVAMERKSQIYLHKVRLQILKRNIGYSAMWAQSCSLPYFPHALELLHLEVEEA  
TSKQIPDAQLPSILDIFIREPVIETIVQCARKTEIALWPYLFMSNAGPKDLFOMCL  
QSEQDTAASTLIILQNLPEVSYSKOYATMLDILALQQRKELADLRLFLKAIDPNE  
IDSPRSNVNWKIAPPVNTQQVQNADAFNMVGLPIARERSFTVTNLSNPKDK  
QASGTPGVAPVTSESSAGAPSVRRSTQRETFICDILTLQHRHQLLQNHKLMIDLY  
MCAYLDLHVLWSQESERAAKLDQFAGALQALHELDLPFPPTAAKDDPAQIRGSL  
RQTGGGSSQTSSEGYFSIATPNGAATSPOLQPSIRREEELQOQPSILPKTRSGS  
QLSFNFRRLYSLPTSEDDLDLIPQKLSIKRLYLLOLFIEANCZDYALVLSILL  
QQAASIRVINGIINSESVHTCRRDLSALKQSLQSFHEGSLYFGFVTLQPHVYLL  
EYIQLSDQPCSQLDAGPCTEQGVDSVTGLQIQNEVEFPVPSQANGQMTVAD  
LNPNHQRLTRIASLESNGNAVVGSSASHSTPTQRLSRQNSRREGCRLM"  
Join(69753. .69861,69916. .69967,70021. .71690,71763. .>72208)  
/gene="Ddx1"  
/note="Nucleotide sequence of the Celera sequence differs

from the published sequence for this transcript."  
/product="CT25986"  
/db\_xref="FLYBASE:FBan0009054"  
/db\_xref="FLYBASE:FBgn0015075"  
<69753. .>72208  
/gene="Ddx1"  
/note="CG9054"  
/map="79E1-79E2"  
/db\_xref="FLYBASE:FBan0009054"  
/db\_xref="FLYBASE:FBgn0015075"  
Join(69846. .69861,69916. .69967,70021. .71690,71763. .72208)  
/gene="Ddx1"  
/note="Ddx1 gene product; Nucleotide sequence of the  
Celera sequence differs from the published sequence for  
this transcript"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009054"  
/db\_xref="FLYBASE:FBgn0015075"  
/protein\_id="AAF51814.1"  
/db\_xref="GI:7296530"  
Query Match 81.7%; Score 18.8; DB 3; Length 295566;  
Best Local Similarity 90.9%; Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gaattctgcttcgtctgttttc 22  
||||| ||||||| |||||||  
Db 31982 GAATTGCTGCTTCTCTGCTTTTC 31961  
RESULT 11  
AX122257  
LOCUS AX122257 1305 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 2173 from Patent Epl108790.  
ACCESSION AX122257  
VERSION AX122257.1 GI:14038973  
KEYWORDS  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 1305)  
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,  
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 2173 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
source location/Qualifiers  
1. .1305  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
BASE COUNT 232 a 298 c 389 g 386 t  
ORIGIN  
Query Match 80.0%; Score 18.4; DB 6; Length 1305;  
Best Local Similarity 95.0%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 aattctgcttcgtctgtttt 21  
||||| ||||||| |||||||  
Db 609 AATTACTGCTTCTGCTTTT 628  
RESULT 12  
AL390862/c  
LOCUS AL390862 112618 bp DNA linear PRI 12-OCT-2001  
DEFINITION Human DNA sequence from clone RP11-30G13 on chromosome 10, complete  
sequence.  
ACCESSION AL390862  
VERSION AL390862.24 GI:16116471  
KEYWORDS HTG.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 112618)  
 Tracey, A.  
 Direct Submission  
 Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clones@requests.sanger.ac.uk  
 On Oct 12, 2001 this sequence version replaced gi:14787516.  
 During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-30G13 is from the library RPci-11.1 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP11-30G13. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true left end of clone RP11-103A2 is at 110619 in this sequence. The true right end of clone RP11-94M14 is at 2000 in this sequence.

FEATURES  
 source  
 1..112618  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-30G13"  
 /clone\_lib="RPci-11.1"  
 9227..9232  
 /note="Single clone region. Assembly confirmed by restriction digest data."

misc\_feature  
 BASE COUNT 36195 a 21831 c 20733 g 33859 t  
 ORIGIN

Query Match 80.0%; Score 18.4; DB 9; Length 112618;  
 Best Local Similarity 95.0%; Pred. NO.4e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctgcttcgtgtttt 21  
 |||||  
 Db 26288 AATTCCTGCTTCTCTGTTT 26269

RESULT 13  
 AC012055/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4, clone RP11-248N22, complete sequence.  
 ACCESSION AC012055  
 VERSION AC012055.9 GI:8671944  
 KEYWORDS HTG.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 171943)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Unpublished  
 Homo sapiens chromosome 4, clone RP11-248N22  
 2 (bases 1 to 171943)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hags, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (19-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 171943)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hags, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (30-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 23, 2000 this sequence version replaced gi:8569149.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L1734  
 Center clone name: 248\_N\_22  
 ----- Location/Qualifiers  
 1..171943  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-248N22"  
 /clone\_lib="RPci-11 Human Male BAC"  
 complement(2..780)  
 repeat\_region

FEATURES  
 source  
 1..171943  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-248N22"  
 /clone\_lib="RPci-11 Human Male BAC"  
 complement(2..780)  
 repeat\_region



repeat\_region /rpt\_family="L1PB1"  
810. .1277  
repeat\_region /rpt\_family="LTR40b"  
1280. .1480  
repeat\_region /rpt\_family="L2"  
complement(2269. .2496)  
repeat\_region /rpt\_family="L2"  
2578. .2710  
repeat\_region /rpt\_family="L2"  
complement(2724. .2800)  
repeat\_region /rpt\_family="L2"  
4277. .4454  
repeat\_region /rpt\_family="MIR"  
4456. .4703  
repeat\_region /rpt\_family="L1MC5"  
complement(5952. .6254)  
repeat\_region /rpt\_family="AluSq"  
6257. .6279  
repeat\_region /rpt\_family="AT\_rich"  
complement(6280. .6568)  
repeat\_region /rpt\_family="AluSc"  
6569. .6585  
repeat\_region /rpt\_family="AT\_rich"  
complement(7177. .7226)  
repeat\_region /rpt\_family="NSTA"  
7617. .7665  
repeat\_region /rpt\_family="AT\_rich"  
complement(8417. .8633)  
repeat\_region /rpt\_family="AluSq"  
10488. .10560  
repeat\_region /rpt\_family="MIR"  
complement(11030. .11134)  
repeat\_region /rpt\_family="MLT1B"  
11135. .11307  
repeat\_region /rpt\_family="MER41B"  
11308. .11536  
repeat\_region /rpt\_family="L1PA7"  
11537. .11912  
repeat\_region /rpt\_family="MER41B"  
complement(11925. .12155)  
repeat\_region /rpt\_family="AluSq"  
12179. .12302  
repeat\_region /rpt\_family="MER41B"  
complement(12312. .12556)  
repeat\_region /rpt\_family="MLT1B"  
complement(12773. .13068)  
repeat\_region /rpt\_family="AluSx"  
13453. .13530  
repeat\_region /rpt\_family="(TATAA)n"  
16473. .16498  
repeat\_region /rpt\_family="(TG)n"  
complement(17063. .17391)  
repeat\_region /rpt\_family="AluSq"  
complement(17608. .17760)  
repeat\_region /rpt\_family="L1M4"  
17910. .17968  
repeat\_region /rpt\_family="MIR"  
18124. .18417  
repeat\_region /rpt\_family="AluSx"  
18634. .18677  
repeat\_region /rpt\_family="AT\_rich"  
19011. .19032  
repeat\_region /rpt\_family="AT\_rich"  
19283. .19412  
repeat\_region /rpt\_family="MIR"  
complement(21374. .21657)  
repeat\_region /rpt\_family="AluSg"  
23426. .23553  
repeat\_region /rpt\_family="MIR"  
23843. .23864  
repeat\_region /rpt\_family="AT\_rich"  
complement(24069. .24159)  
repeat\_region /rpt\_family="MLT1K"

repeat\_region complement(24164. .24289)  
repeat\_region /rpt\_family="AluJb"  
25593. .25631  
repeat\_region /rpt\_family="(TTTA)n"  
complement(25849. .26021)  
repeat\_region /rpt\_family="MIR"  
26807. .27161  
repeat\_region /rpt\_family="THE1B"  
28665. .29002  
repeat\_region /rpt\_family="MLT1A1"  
31978. .32056  
repeat\_region /rpt\_family="L1MC4"  
32109. .32328  
repeat\_region /rpt\_family="L1MC4"  
33299. .33452  
repeat\_region /rpt\_family="(TATAA)n"  
33513. .33577  
repeat\_region /rpt\_family="AT\_rich"  
34264. .34297  
repeat\_region /rpt\_family="(TTTTTG)n"  
35484. .35507  
repeat\_region /rpt\_family="AT\_rich"  
36140. .36186  
repeat\_region /rpt\_family="AT\_rich"  
36284. .36312  
repeat\_region /rpt\_family="AT\_rich"  
37765. .37791  
repeat\_region /rpt\_family="AT\_rich"  
37856. .37878  
repeat\_region /rpt\_family="AT\_rich"  
38007. .38035  
repeat\_region /rpt\_family="AT\_rich"  
complement(40978. .41140)  
repeat\_region /rpt\_family="MER103"  
41297. .41330  
repeat\_region /rpt\_family="(TAGA)n"  
complement(41789. .41848)  
repeat\_region /rpt\_family="L2"  
complement(41803. .41938)  
repeat\_region /rpt\_family="MIR"  
43710. .43851  
repeat\_region /rpt\_family="MIR"  
43958. .44767  
repeat\_region /rpt\_family="L1ME3"  
45733. .45761  
repeat\_region /rpt\_family="AT\_rich"  
complement(47019. .47316)  
repeat\_region /rpt\_family="AluY"  
48229. .48276

Query Match 80.0%; Score 18.4; DB 9; Length 171943;  
Best Local Similarity 95.0%; Pred. NO. 3.8e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctgcttcgctgtttt 21  
||||| |||||||  
Db 170336 AATTCTACTTCGCTGTTT 170317

RESULT 14  
AC106895/c

LOCUS AC106895 Homo sapiens chromosome 4 clone RP11-161D15, WORKING DRAFT  
DEFINITION SEQUENCE, 2 unordered pieces.

ACCESSION AC106895  
VERSION AC106895.3 GI:18308813  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 179667)  
AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 179667)  
AUTHORS Waterston, R. H.  
TITLE Direct Submission  
JOURNAL Submitted (13-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Jan 24, 2002 this sequence version replaced gi:18141525.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@wustl.wustl.edu  
----- Project Information -----  
Center project name: H\_NH0161D15  
----- Summary Statistics -----  
Sequencing vector: M13: 0%  
Chemistry: Dye-terminator ET; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 178624 bases at least Q40  
Consensus quality: 178982 bases at least Q30  
Consensus quality: 179266 bases at least Q20  
Insert size: 173000; agarose-fp  
Insert size: 179567; sum-of-contigs  
Quality coverage: 9.03 in Q20 bases; agarose-fp  
Quality coverage: 8.15 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 65346: contig of 65346 bp in length  
\* 65347 65446: gap of unknown length  
\* 65447 179667: contig of 114221 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..179667  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="RP11-161D15"  
misc\_feature  
1..65346  
/note="assembly\_name:Contig98"  
misc\_feature  
65447..179667  
/note="assembly\_name:Contig99"  
clone\_end:SP6  
vector\_side:right  
BASE COUNT 57454 a 33427 c 31929 g 56757 t 100 others  
ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 179667;  
Best Local Similarity 95.0%; Pred. No. 3.8e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattcctgcttcgctgtttt 21  
||||| |||||||  
Db 44493 AATTCTACTTCGCTGTTT 44474

RESULT 15  
AX127149/c 349980 bp DNA linear PAT 11-MAY-2001  
LOCUS AX127149  
DEFINITION Sequence 7065 from Patent EP1108790.  
ACCESSION AX127149 AX114121

VERSION AX127149.1 GI:14041137  
KEYWORDS Corynebacterium glutamicum.  
SOURCE Corynebacterium glutamicum  
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
REFERENCE 1 (bases 1 to 349980)  
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7065 20-JUN-2001;  
FEATURES Location/Qualifiers  
source  
1..349980  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3.309.400) split in 11, seq 7065  
1.800.001 2.149.980"  
BASE COUNT 86896 a 98023 c 80939 g 84122 t  
ORIGIN

Query Match 80.0%; Score 18.4; DB 6; Length 349980;  
Best Local Similarity 95.0%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattcctgcttcgctgtttt 21  
||||| |||||||  
Db 291447 AATTACTGCTTCGCTGTTT 291428

Search completed: July 25, 2002, 05:39:24  
Job time: 6717 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:44:43 ; Search time 378.07 Seconds  
(without alignments)  
104.449 Million cell updates/sec

Title: US-09-235-416-4

Perfect score: 23

Sequence: 1 gaattctgcttcgttttca 23

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	AAH87658	Kinesin motor prot
C 2	23	100.0	2352	AAH87656	Thermomyces lanugi
C 3	18.8	81.7	466	AAH16520	Human gene express
C 4	18.4	80.0	1305	AAH67138	C glutamicum codin
C 5	18.4	80.0	349980	AAH68530	C glutamicum codin
C 6	18.2	79.1	705	AAH56002	Salmonella typhi D
C 7	17.8	77.4	2408	AAH199592	Mouse ischaemic co
C 8	17.4	75.7	2288	AAD08438	Human secreted pro
C 9	17.4	75.7	8734	ABA09012	Human TRAP240 homo

# ALIGNMENTS

RESULT 1

AAH87658

ID AAX87658 standard; DNA; 23 BP.

XX AC AAX87658;

XX DT 26-OCT-1999 (first entry)

XX DE Kinesin motor protein TL-gamma gene PCR primer.

XX TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;

XX KW neurodegenerative disease; Alzheimer's disease;

XX KW Parkinson's disease; Huntington's disease;

XX KW anyotrophic lateral sclerosis; PCR; primer; ss.

XX OS Synthetic.

XX OS Thermomyces lanuginosus.

XX PN WO9937659-A1.

XX PD 29-JUL-1999.

XX PF 22-JAN-1999; 99WO-US01355.

XX PR 23-JAN-1998; 98US-0072361.

XX PA (REGC ) UNIV CALIFORNIA.

XX XX Goldstein LSB, Sakowicz R;

XX XX WPI; 1999-493950/41.

C 10	17.4	75.7	8734	22	AAI59594	Human polynucleoti
C 11	17.4	75.7	8761	22	AAI57808	Human polynucleoti
C 12	17.4	75.7	8785	23	AAH85661	DNA encoding novel
C 13	17.2	74.8	4016	22	AAK77376	Human immune/haema
C 14	17.2	74.8	4762	22	AAK77377	Human immune/haema
C 15	16.8	73.0	360	19	AAH14159	H. pylori GHPO 491
C 16	16.8	73.0	4078	22	AAH16633	Human cDNA sequenc
C 17	16.8	73.0	4715	22	ABAI6034	Human nervous syst
C 18	16.8	73.0	4715	22	ABAI6035	Human nervous syst
C 19	16.8	73.0	127197	22	AAI61370	Soybean 515002 reg
C 20	16.6	72.2	212	21	AAH03476	Human secreted pro
C 21	16.6	72.2	216	20	AAH03476	Human secreted pro
C 22	16.6	72.2	393	20	AAH20106	Enterococcus faeca
C 23	16.6	72.2	538	22	AAH40008	Genomic sequence #
C 24	16.6	72.2	538	22	AAH40008	Human digestive sy
C 25	16.6	72.2	1965	23	AAH66737	DNA encoding novel
C 26	16.6	72.2	4758	18	AAH96561	Rat neurodapl 1 gen
C 27	16.6	72.2	4778	18	AAH96562	Human neurodapl 1 g
C 28	16.6	72.2	20561	20	AAH13229	Enterococcus faeca
C 29	16.6	72.2	58708	22	AAH64739	Human immune/haema
C 30	16.6	72.2	72604	20	AAH10752	Genomic sequence o
C 31	16.6	72.2	119950	20	AAH90201	Human yes1 gene.
C 32	16.6	72.2	236303	22	AAH11614	Human genomic DNA
C 33	16.4	71.3	31	20	AAH39026	Human genomic DNA
C 34	16.2	70.4	196	21	AAH43132	Human secreted exp
C 35	16.2	70.4	198	21	AAH08506	Human secreted pro
C 36	16.2	70.4	406	21	AAH21446	Human secreted pro
C 37	16.2	70.4	412	21	AAH43726	Mouse secreted exp
C 38	16.2	70.4	485	22	AAH24961	Human ovarian PCR-
C 39	16.2	70.4	485	22	AAH83606	Human ovarian tumo
C 40	16.2	70.4	504	20	AAH88728	EST clone HJ181.
C 41	16.2	70.4	1667	21	AAH62532	Human secreted pro
C 42	16.2	70.4	1667	21	AAH22706	Human secreted pro
C 43	16.2	70.4	2345	19	AAH71209	DNA encoding ester
C 44	16.2	70.4	2405	19	AAH71215	DNA encoding ester
C 45	16.2	70.4	3155	24	AAH94847	Human DNA sequence



XX The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AA12532 to AA21779. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AA12532 to AA21779. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct  
CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX SQ Sequence 466 BP; 185 A; 65 C; 97 G; 116 T; 3 other;

Query Match 81.7%; Score 18.8; DB 20; Length 466;  
Best Local Similarity 90.9%; Pred. No. 41;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaattctgtctgtctgttttc 22  
|||||  
Db 434 GAATTCCTGCTCCATGTTTC 413

RESULT 4  
AAH67138  
ID AAH67138 standard; DNA; 1305 BP.  
AC AAH67138;  
DT 26-SEP-2001 (first entry)  
DE C glutamicum coding sequence fragment SEQ ID NO: 2173.  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX EPI108790-A2.  
XX 20-JUN-2001.  
XX 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI: 2001-376931/40.  
DR P-FSDB; AAG91919.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 8; SEQ ID NO: 2173; 246pp + Sequence Listing; English.  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX SQ Sequence 1305 BP; 232 A; 298 C; 389 G; 386 T; 0 other;

Query Match 80.0%; Score 18.4; DB 22; Length 1305;  
Best Local Similarity 95.0%; Pred. No. 67;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctctgtctgtgtttt 21  
|||||  
Db 609 aattactgtctgtgtttt 628

RESULT 5  
AAH68530/c  
ID AAH68530 standard; DNA; 349980 BP.  
AC AAH68530;  
DT 26-SEP-2001 (first entry)  
DE C glutamicum coding sequence fragment SEQ ID NO: 7065.  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX EPI108790-A2.  
XX 20-JUN-2001.  
XX 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI: 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

XX Disclosure: SEQ ID NO: 7065; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived









XX 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 11.  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
XX P-PSDB; AAM38652.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 11; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence 8761 BP; 2551 A; 1778 C; 1784 G; 2648 T; 0 other;

Query Match 75.7%; Score 17.4; DB 22; Length 8761;  
Best Local Similarity 94.7%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 gaattctcgttcgctgtt 19  
|||||  
DB 3144 GAATTCCTGCTCGCTGT 3126

RESULT 12

AAS85661/c  
ID AAS85661 standard; cDNA; 8785 BP.  
XX  
XX AAS85661;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #21465.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX P-PSDB; ABG21474.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID NO 21465; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 8785 BP; 2531 A; 1811 C; 1806 G; 2637 T; 0 other;

Query Match 75.7%; Score 17.4; DB 23; Length 8785;  
Best Local Similarity 94.7%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 gaattctcgttcgctgtt 19  
|||||  
DB 3104 GAATTCCTGCTCGCTGT 3086

RESULT 13  
AAK77376/c

ID XX AAK77376 standard; DNA; 4016 BP.  
AC AAK77376;  
XX DT 07-NOV-2001 (first entry)  
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32188.  
DE DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX OS Homo sapiens.  
XX PN WO200157182-A2.  
XX PD 09-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.

```
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254997.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX DR WPI; 2001-483426/52.
XX
XX PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX PS Disclosure; SEQ ID NO 32188; 3071pp + Sequence Listing; English.
XX
XX CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 4016 BP; 1298 A; 763 C; 795 G; 1160 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4016;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aattcctgcttcgttttca 23
    ||||| ||||| ||||| |||||
Db 2175 AATTTCGCTTACCTGTTTCA 2154

RESULT 14
AAK77377/c
ID AAK77377 standard; DNA; 4762 BP.
XX
XX AC AAK77377;
XX
XX XX
XX DT 07-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32189.
XX
XX KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
XX PN HQ200157182-A2.
XX
XX XX
XX PD 09-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01354.
XX
XX
```

```
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
```

XX	02-OCT-2000;	2000US-0237037.	
PS	02-OCT-2000;	2000US-0237038.	
XX	02-OCT-2000;	2000US-0237039.	
CC	02-OCT-2000;	2000US-0237040.	
CC	13-OCT-2000;	2000US-0239935.	
CC	13-OCT-2000;	2000US-0239937.	
CC	20-OCT-2000;	2000US-0240960.	
CC	20-OCT-2000;	2000US-0241221.	
CC	20-OCT-2000;	2000US-0241785.	
CC	20-OCT-2000;	2000US-0241786.	
CC	20-OCT-2000;	2000US-0241787.	
CC	20-OCT-2000;	2000US-0241808.	
CC	20-OCT-2000;	2000US-0241809.	
CC	20-OCT-2000;	2000US-0241826.	
CC	01-NOV-2000;	2000US-0244617.	
CC	08-NOV-2000;	2000US-0246474.	
CC	08-NOV-2000;	2000US-0246475.	
CC	08-NOV-2000;	2000US-0246476.	
CC	08-NOV-2000;	2000US-0246477.	
CC	08-NOV-2000;	2000US-0246478.	
CC	08-NOV-2000;	2000US-0246523.	
CC	08-NOV-2000;	2000US-0246524.	
CC	08-NOV-2000;	2000US-0246525.	
CC	08-NOV-2000;	2000US-0246526.	
CC	08-NOV-2000;	2000US-0246527.	
CC	08-NOV-2000;	2000US-0246528.	
CC	08-NOV-2000;	2000US-0246532.	
CC	08-NOV-2000;	2000US-0246609.	
CC	08-NOV-2000;	2000US-0246610.	
CC	08-NOV-2000;	2000US-0246611.	
CC	08-NOV-2000;	2000US-0246613.	
CC	17-NOV-2000;	2000US-0249207.	
CC	17-NOV-2000;	2000US-0249208.	
CC	17-NOV-2000;	2000US-0249209.	
CC	17-NOV-2000;	2000US-0249210.	
CC	17-NOV-2000;	2000US-0249211.	
CC	17-NOV-2000;	2000US-0249212.	
CC	17-NOV-2000;	2000US-0249213.	
CC	17-NOV-2000;	2000US-0249214.	
CC	17-NOV-2000;	2000US-0249215.	
CC	17-NOV-2000;	2000US-0249216.	
CC	17-NOV-2000;	2000US-0249217.	
CC	17-NOV-2000;	2000US-0249218.	
CC	17-NOV-2000;	2000US-0249218.	
CC	17-NOV-2000;	2000US-0249244.	
CC	17-NOV-2000;	2000US-0249245.	
CC	17-NOV-2000;	2000US-0249245.	
CC	17-NOV-2000;	2000US-0249264.	
CC	17-NOV-2000;	2000US-0249265.	
CC	17-NOV-2000;	2000US-0249297.	
CC	17-NOV-2000;	2000US-0249299.	
CC	17-NOV-2000;	2000US-0249300.	
CC	01-DEC-2000;	2000US-0250160.	
CC	01-DEC-2000;	2000US-0250391.	
CC	05-DEC-2000;	2000US-0251030.	
CC	05-DEC-2000;	2000US-0251988.	
CC	05-DEC-2000;	2000US-0256719.	
CC	06-DEC-2000;	2000US-0251479.	
CC	08-DEC-2000;	2000US-0251856.	
CC	08-DEC-2000;	2000US-0251868.	
CC	08-DEC-2000;	2000US-0251869.	
CC	08-DEC-2000;	2000US-0251989.	
CC	08-DEC-2000;	2000US-0251990.	
CC	11-DEC-2000;	2000US-0254097.	
CC	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-483426/52.		
XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
XX	useful for preventing, diagnosing and/or treating cancers and		
XX	metastasis -		

XX	Disclosure; SEQ ID NO 32189; 3071pp + Sequence Listing; English.			
XX	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)			
CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic			
CC	activity, and can be used in gene therapy and vaccine production. (I)			
CC	proteins and polynucleotides may be used in the prevention, diagnosis and			
CC	treatment of diseases associated with inappropriate (I) expression. For			
CC	example, they may be used to treat disorders associated with decreased			
CC	expression by rectifying mutations or deletions in a patient's genome			
CC	that affect the activity of (I) by expressing inactive proteins or to			
CC	supplement the patients own production of (I). Additionally, (I)			
CC	polynucleotides may be used to produce the secreted (I), by inserting			
CC	the nucleic acids into a host cell and culturing the cell to express the			
CC	protein. (I) proteins and polynucleotides may be used to prevent,			
CC	diagnose and treat immune/haematopoietic-related diseases, especially			
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703			
CC	to AAK87694 represent human immune/haematopoietic antigen genomic			
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169			
CC	represent sequences used in the exemplification of the present invention			
XX	Sequence 4762 BP; 1540 A; 898 C; 947 G; 1377 T; 0 other;			
SQ				
	Query Match 74.8%; Score 17.2; DB 22; Length 4762;			
	Best Local Similarity 86.4%; Pred. No. 2.5e+02;			
	Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	2 aattctgcttgcgtgtttca 23			
Db	2923 AATTTCGCTTACCTGTTTCA 2902			
RESULT 15				
AAK14159/C				
ID	AAK14159 standard; DNA; 360 BP.			
XX	AC AAK14159;			
XX	31-MAR-1999 (first entry)			
DE	H. pylori GHPO 491 gene.			
XX	GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;			
KW	peptic ulcer disease; ss.			
OS	Helicobacter pylori.			
PH	Key Location/Qualifiers			
FT	CDS 82..297			
FT	/*tag= a			
XX	WO9843478-A1.			
XX	08-OCT-1998.			
XX	01-APR-1998; 98WO-US06371.			
XX	29-JUL-1997; 97US-0902615.			
PR	01-APR-1997; 97US-0833457.			
PR	24-JUN-1997; 97US-0881227.			
XX	(HUMA-) HUMAN GENOME SCI INC.			
PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.			
XX	Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;			
XX	WPI; 1998-542293/46.			
DR	P-PSDB; AAW98440.			
XX	New isolated Helicobacter polynucleotides - used to develop products			
PT	for the diagnosis, prevention and treatment of Helicobacter			
PT	infections and gastrointestinal diseases			

XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

Claim 1; Page 756; 2054pp; English.

This sequence represents a polynucleotide of the invention. It was isolated from *Helicobacter pylori* and encodes a *H. pylori* GHPO protein. The polypeptides can be used for preventing or treating *Helicobacter* infections, and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

Sequence 360 BP; 160 A; 64 C; 71 G; 65 T; 0 other;

Query Match 73.0%; Score 16.8; DB 19; Length 360;  
Best Local Similarity 90.0%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ttcttgcttcgctgtttcca 23  
|| |||||  
Db 169 TTAGTCTTCGCTGTTTCA 150

Search completed: July 25, 2002, 05:44:57  
Job time: 6365 sec

**THIS PAGE BLANK (USPTO)**



Thu Jul 25 08:38:13 2002

```
;
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2345
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E006
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(1689)
; US-09-058-260-15
;
Query Match          70.4%; Score 16.2; DB 4; Length 2345;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 gaattctgcttcgctgtttt 21
      ||||| ||| | ||||| |||||
Db      2124 gaatccctccatcgctgtttt 2144

RESULT      3
US-09-058-260-27
; Sequence 27, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Aikens, John
; APPLICANT: Fonstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
```

```
;
; LENGTH: 2405
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)..(1766)
; US-09-058-260-27
;
Query Match          70.4%; Score 16.2; DB 4; Length 2405;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 gaattctgcttcgctgtttt 21
      ||||| ||| | ||||| |||||
Db      3125 gaatccctccatcgctgtttt 3145

RESULT      5
US-09-058-260-1
; Sequence 1, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Aikens, John
; APPLICANT: Fonstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E001
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1182)..(2690)
; US-09-058-260-1
;
Query Match          70.4%; Score 16.2; DB 4; Length 3513;
Best Local Similarity 85.7%; Pred. No. 86;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 gaattctgcttcgctgtttt 21
      ||||| ||| | ||||| |||||
Db      3125 gaatccctccatcgctgtttt 3145

RESULT      5
```



US-08-781-802-9  
; Sequence 9, Application US/08781802  
; Patent No. 5969121  
; GENERAL INFORMATION:  
; APPLICANT: ALLEN, Larry  
; APPLICANT: AIKENS, John  
; APPLICANT: FOSTEIN, Michael  
; APPLICANT: VONSTEIN, Veronika  
; APPLICANT: DEMIRJIAN, David  
; APPLICANT: CASADABAN, Malcolm  
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 S. Wacker Drive 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,802  
; FILING DATE: 10-JAN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/694,078  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/019,580  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/009,704  
; FILING DATE: 11-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,995  
; FILING DATE: 01-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 95,963-E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1397..2905  
; OTHER INFORMATION: /note= "E019 sequence of longest  
; OTHER INFORMATION: open reading frame; upstream untranslated region not exact"  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1397..2905  
US-08-781-802-9  
Query Match 70.4%; Score 16.2; DB 2; Length 3545;  
Best Local Similarity 85.7%; Pred. No. 86;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgtgtttt 21

Db 3340 GAATCCCTCCATCGTGTGTTT 3360  
||||| ||| | ||||| |||||  
RESULT 6  
US-08-781-802-1  
; Sequence 1, Application US/08781802  
; Patent No. 5969121  
; GENERAL INFORMATION:  
; APPLICANT: ALLEN, Larry  
; APPLICANT: AIKENS, John  
; APPLICANT: FOSTEIN, Michael  
; APPLICANT: VONSTEIN, Veronika  
; APPLICANT: DEMIRJIAN, David  
; APPLICANT: CASADABAN, Malcolm  
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 S. Wacker Drive 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60605  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,802  
; FILING DATE: 10-JAN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/694,078  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/019,580  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/009,704  
; FILING DATE: 11-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,995  
; FILING DATE: 01-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 95,963-E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4315 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1182..2690  
; OTHER INFORMATION: /note= "E001, longest open reading  
; OTHER INFORMATION: frame; other possible start codons at ATG/met5; GTG/val8;  
; OTHER INFORMATION: GTG/val10; TTG/leu17"  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1182..2690  
US-08-781-802-1

Query Match 70.4%; Score 16.2; DB 2; Length 4315;

Best Local Similarity 85.7%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 3; Indels 3; Gaps 0;  
Qy 1 gaattcgttcgctgtttt 21  
||||| ||| | |||||  
Db 3125 GAATCCCTCCATCGCTGTTT 3145

RESULT 7  
US-08-694-078-1  
; Sequence 1, Application US/08694078  
; Patent No. 6218163  
; GENERAL INFORMATION:  
; APPLICANT: ALLEN, Larry  
; APPLICANT: AIKENS, John  
; APPLICANT: FONSTEIN, Michael  
; APPLICANT: VONSTEIN, Veronika  
; APPLICANT: DEMIRJIAN, David  
; APPLICANT: CASADABAN, Malcolm  
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff, Ltd.  
; STREET: 300 S. Wacker Drive 7th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,078  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/019,580  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/009,704  
; FILING DATE: 10-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,995  
; FILING DATE: 07-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 95,963-C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4315 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1182..2690  
; OTHER INFORMATION: /note= "E001, longest open reading  
; frame; other possible start codons at ATG/met5; GTG/val8;  
; OTHER INFORMATION: GTG/val10; TTG/leu17"  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1182..2690  
; US-08-694-078-1

Query Match 70.4%; Score 16.2; DB 4; Length 4315;

Best Local Similarity 85.7%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 3; Indels 3; Gaps 0;  
Qy 1 gaattcgttcgctgtttt 21  
||||| ||| | |||||  
Db 3125 GAATCCCTCCATCGCTGTTT 3145

RESULT 8  
US-09-008-979A-8  
; Sequence 8, Application US/09008979A  
; Patent No. 6080914  
; GENERAL INFORMATION:  
; APPLICANT: Conner, Timothy W.  
; TITLE OF INVENTION: Strawberry Promoters and Genes  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/008,979A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kammerer, Patricia A.  
; REGISTRATION NUMBER: 29,775  
; REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 787-1438  
; TELEFAX: (713) 789-2679  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 752 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-09-008-979A-8

Query Match 68.7%; Score 15.8; DB 3; Length 752;  
Best Local Similarity 89.5%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ttctcgttcgctgttttc 22  
||| ||||| |||||  
Db 40 TTTCGTCTCTCTGTTTTC 58

RESULT 9  
US-09-460-618-8  
; Sequence 8, Application US/09460618  
; Patent No. 6235482  
; GENERAL INFORMATION:  
; APPLICANT: Conner, Timothy W.  
; TITLE OF INVENTION: Strawberry Promoters and Genes  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/460,618
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,979
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1438
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-460-618-8

```

```

Query Match      68.7%; Score 15.8; DB 4; Length 752;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4  ttctgcttcgctgtttc 22
          || ||||| |||||
Db      40  ttTCTGCTTCTCTGTTTC 58

```

RESULT 10  
US-08-474-140-15  
Sequence 15, Application US/08474140  
Patent No. 5721127  
GENERAL INFORMATION:  
APPLICANT: DEWEER, PHILIPPE  
APPLICANT: AMORY, ANTOINE  
TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH  
PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS  
TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.  
STREET: 2000 K Street, N.W., Suite 200  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,140  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilhem F. Gadiano, Esq.  
REGISTRATION NUMBER: 37,136  
REFERENCE/DOCKET NUMBER: 4121-41  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 429-0625  
TELEFAX: (202) 293-1850  
TELEX: 650 383-5605  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
US-08-474-140-15

```

Query Match	67.8%	Score 15.6;	DB 1;	Length 431;
Best Local Similarity	81.8%;	Pred. No. 1.2e+02;		
Matches 18;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	2	aattcctgcgttcggtttcca	23	
Db	66	ATTTCGCGCTTCGGTGTTC	87	

RESULT 11  
US-08-477-630-15  
; Sequence 15, Application US/08477630  
; Patent No. 5721128  
; GENERAL INFORMATION:  
; APPLICANT: DEWEER, PHILIPPE  
; APPLICANT: AMORY, ANTOINE  
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH  
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS  
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.  
; STREET: 2000 K Street, N.W., Suite 200  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,630  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wilhem F. Gadiano, Esq.  
; REGISTRATION NUMBER: 37,136  
; REFERENCE/DOCKET NUMBER: 4121-42  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 429-0625  
; TELEFAX: (202) 293-1850  
; TELEX: 650 383-5605  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 431 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: nucleic acid  
; US-08-477-630-15

```

Query Match      67.8%; Score 15.6; DB 1; Length 431;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY      2 aattcctgcttcgctgtttca 23
      | | | | | | | | | | | | |
Db      66 ATTTCGCGCTTCGCTGTTTCA 87

```

RESULT 12  
US-08-472-293-15

; Sequence 15, Application US/08472293  
; Patent No. 5731174

## GENERAL INFORMATION:

APPLICANT: DEWEER, PHILIPPE

APPLICANT: AMORY, ANTOINE

TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH

TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS

TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: WILLIAN BRINKS HOFER GILSON &amp; LIONE, P.C.

STREET: 2000 K Street, N.W., Suite 200

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/472,293

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wilhem F. Gadiano, Esq.

REGISTRATION NUMBER: 37,136

REFERENCE/DOCKET NUMBER: 4121-44

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 429-0625

TELEFAX: (202) 293-1850

TELEX: 650 383-5605

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 431 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: nucleic acid

US-08-472-293-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;

Best Local Similarity 81.8%; Pred. No. 1.2e+02;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 aattcgtctgcgtgtttca 23

Db 66 ATTTCGGCTTCGGTGTTC 87

## RESULT 13

US-08-474-545-15

; Sequence 15, Application US/08474545

; Patent No. 5736375

## GENERAL INFORMATION:

APPLICANT: DEWEER, PHILIPPE

APPLICANT: AMORY, ANTOINE

TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH

TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS

TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: WILLIAN BRINKS HOFER GILSON &amp; LIONE, P.C.

STREET: 2000 K Street, N.W., Suite 200

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,545

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wilhem F. Gadiano, Esq.

REGISTRATION NUMBER: 37,136

REFERENCE/DOCKET NUMBER: 4121-43

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 429-0625

TELEFAX: (202) 293-1850

TELEX: 650 383-5605

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 431 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: nucleic acid

US-08-474-545-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;

Best Local Similarity 81.8%; Pred. No. 1.2e+02;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 aattcgtctgcgtgtttca 23

Db 66 ATTTCGGCTTCGGTGTTC 87

## RESULT 14

US-08-478-341-15

; Sequence 15, Application US/08478341

; Patent No. 5817498

## GENERAL INFORMATION:

APPLICANT: DEWEER, PHILIPPE

APPLICANT: AMORY, ANTOINE

TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH

TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS

TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: WILLIAN BRINKS HOFER GILSON &amp; LIONE, P.C.

STREET: 2000 K Street, N.W., Suite 200

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/478,341

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wilhem F. Gadiano, Esq.

REGISTRATION NUMBER: 37,136

REFERENCE/DOCKET NUMBER: 4121-45

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 429-0625

TELEFAX: (202) 293-1850

TELEX: 650 383-5605

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 431 base pairs

TYPE: nucleic acid

STRANDEDNESS: single



**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USP1U)**



IntelliGenetics

## FastDB - Fast Pairwise Comparison of Sequences Release 5.4

Results file seq2-1327-1803-aa.res made by bobryen on Tue 23 Jul 102 15:29:45-PDT.

Query sequence being compared:	US-09-235-416-2 (1327-1803)
Number of sequences searched:	1
Number of scores above cutoff:	1

Results of the initial comparison of US-09-235-416-2 (1327-1803) with:  
File : US0235416A.pep

Item	Score
N	18
D	35
M	53
B	71
E	88
R	106
O	124
F	141
S	159
E	159
Q	159
U	159
E	159
N	159
C	159
E	159
S	159

## PARAMETERS

Similarity matrix	PAM-150	K-tuple	1
Threshold level of sim.	16%		
Translation Frame	6		
Mismatch penalty	1	Joining penalty	20
Gap penalty	5.00	Window size	477
Gap size penalty	0.05		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	159	0	0.00

Times:	CPU	Total Elapsed
	00:00:00.00	00:00:00.00

Number of residues:	784
Number of sequences searched:	1
Number of scores above cutoff:	1

The scores below are sorted by initial score. Significance is calculated based on initial s

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Init. Opt.
1. US-09-235-416-1	Sequence 1, Application US	784	159	0.00

1. US-09-235-416-2 (1327-1803)

US-09-235-416-1 Sequence 1, Application US/09235416A

Initial Score	=	159	Optimized Score	=	159	Significance	=	0.00
Residue Identity	=	100%	Matches	=	159	Mismatches	=	0
Gaps	=		Conservative Substitutions	=			=	0
Translation Frame	=	1						

[illegible]

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**



AAGCTGGTCTCGCGATCATTTCCACAGAAATCAAAATCTCATGGATGAGGGCAACAAGACAGAAACGGTTCGG	580	590	600	610	620	630	640		
GCCACAAACATGAACGAGACATCCAGTCGATCCACGCCGCTTTCACCTTGACCTTGACCGAAAAGTGGCAT	650	660	670	680	690	700	710	720	
GCCACAAACATGAACGAGACATCCAGTCGATCCACGCCGCTTTCACCTTGACCTTGACCGAAAAGTGGCAT	650	660	670	680	690	700	710	720	
GATGAAGAGACCAAAATGGACACAGAGAAGTTGCGAGAGATCAGTCTGGTAGATTTTGGCGGGTCTCGAGCGA	730	740	750	760	770	780	790		
GATGAAGAGACCAAAATGGACACAGAGAAGTTGCGAGAGATCAGTCTGGTAGATTTTGGCGGGTCTCGAGCGA	730	740	750	760	770	780	790		
GCAACGCTCCACGGAGCTACTGGAGCGCGACTGAAGCGGTGCAGAGATCAACCGTCTACTTTCGACCCCTA	800	810	820	830	840	850	860		
GCAACGCTCCACGGAGCTACTGGAGCGCGACTGAAGCGGTGCAGAGATCAACCGTCTACTTTCGACCCCTA	800	810	820	830	840	850	860		
GGTCGCTGTGATTGCAGCGCTAGCGGATATGTCGTCGGGAAAAACAGAAGAAGATCAGTTAGTACCTTACCGA	870	880	890	900	910	920	930		
GGTCGCTGTGATTGCAGCGCTAGCGGATATGTCGTCGGGAAAAACAGAAGAAGATCAGTTAGTACCTTACCGA	870	880	890	900	910	920	930		
GATTTCGGTACTGACGTGGCTTCTCAAGGACTTCCPTGGGAGCACTCGATGACCGCCATCATTTCCGCCCAAT	940	950	960	970	980	990	1000		
GATTTCGGTACTGACGTGGCTTCTCAAGGACTTCCPTGGGAGCACTCGATGACCGCCATCATTTCCGCCCAAT	940	950	960	970	980	990	1000		
TCGCGCTGCTGATATTAACCTTGAAGAGACTTCTCAGTACCCTTCGATATCGGACTCTCGCGAAG	1010	1020	1030	1040	1050	1060	x		
TCGCGCTGCTGATATTAACCTTGAAGAGACTTCTCAGTACCCTTCGATATCGGACTCTCGCGAAG	1010	1020	1030	1040	1050	1060	1070	1080	
AACACACGACGT	1090								

2. US-09-235-416-2 (1-1071)  
US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score	=	21	Optimized Score	=	21	Significance	=	-0.39
Residue Identity	=	100%	Matches	=	21	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=	0			

X                    10                    X                    30                    40  
ATGTCGGCGGTGGAATATCAAGGTGTGTGTCGGGTACG  
|||||  
ATGTCGGCGGTGGAATATC  
X                    10                    20                    X

3. US-09-235-416-2 (1-1071)  
US-09-235-416-6 Sequence 6, Application US/09235416A

Initial Score	=	17	Optimized Score	=	17	Significance	=	-0.40
Residue Identity	=	40%	Matches	=	17	Mismatches	=	13
Gaps	=	0	Conservative Substitutions	=	0		=	0

800 810 820 830 840 X 850 860  
CGTCCACGGAGCTACTGGAGCGGCATGAGGAGGTGCGAGAGATCAACCGCTCATTTCGACCCTAGG

||||| | ||||| |  
GGCGGAATTTCTCDGANCCDGCVARRTCNAC  
X     10     20     30

4. US-09-235-416-2 (1-1071)  
US-09-235-416-5 Sequence 5, Application US/09235416A

[illegible]

5. US-09-235-416-2 (1-1071)  
US-09-235-416-7 Sequence 7, Application US/09235416A

[illegible]

6. US-09-235-416-2 (1-1071)  
US-09-235-416-4 Sequence 4, Application US/09235416A

```
Initial Score = 10 Optimized Score = 11 Significance = -0.42
Residue Identity = 58% Matches = 14 Mismatches = 9
Gaps = 1 Conservative Substitutions = 0
```

4. US-09-235-416-2 (1-1071)

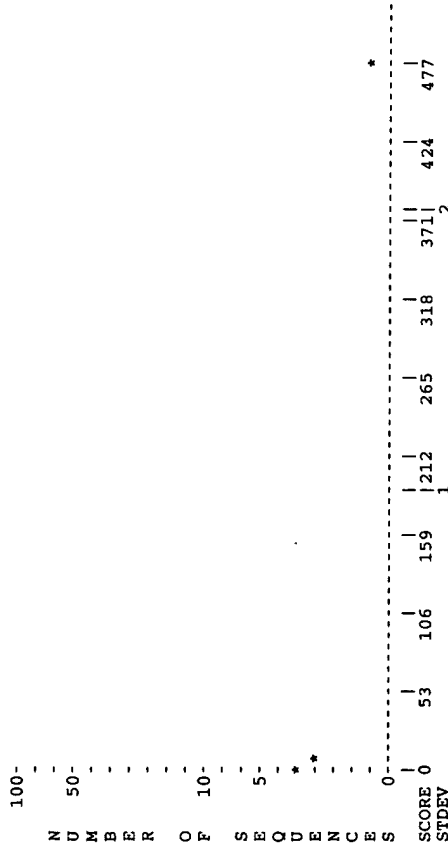
> O <  
OI IO IntelliGenetics  
> O <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file us-09-235-416-2-1327-1803.res made by bobryen on Tue 23 Jul 102 15:24:40-PM

Query sequence being compared: US-09-235-416-2 (1327-1803)  
Number of sequences searched: 6  
Number of scores above cutoff: 6

Results of the initial comparison of US-09-235-416-2 (1327-1803) with:  
File : US09235416A.seq



The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
US-09-235-416-2 (1327-1803)		477	477	2.04

1. US-09-235-416-2 Sequence 2, Application US 2352 477 477 2.04 0

The list of other best scores is:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
2. US-09-235-416-7 Sequence 7, Application US 30	11	14	-0.39	0
3. US-09-235-416-6 Sequence 6, Application US 30	8	15	-0.40	0
4. US-09-235-416-3 Sequence 3, Application US 21	6	9	-0.41	0
5. US-09-235-416-4 Sequence 4, Application US 23	5	12	-0.42	0

1. US-09-235-416-2 (1327-1803)  
US-09-235-416-2 Sequence 2, Application US/09235416A

Initial Score = 477 Optimized Score = 477 Significance = 2.04  
Residue Identity = 100% Matches = 477 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

60 70 80 90 100 110 120  
130 140 150 160 170 180 190  
200 210 220 230 240 250 260  
270 280 290 300 310 320 330 340  
350 360 370 380 390 400 410  
420 430 440 450 460 470 480 490  
500 510 520 530 540 550 560 570 580 590  
600 610 620 630 640 650 660 670 680 690  
700 710 720 730 740 750 760 770 780 790  
800 810 820 830 840 850 860 870 880 890  
900 910 920 930 940 950 960 970 980 990  
1000 1010 1020 1030 1040 1050 1060 1070 1080 1090  
1100 1110 1120 1130 1140 1150 1160 1170 1180 1190  
1200 1210 1220 1230 1240 1250 1260 1270 1280 1290  
1300 1310 1320 1330 1340 1350 1360 1370 1380 1390  
1400 1410 1420 1430 1440 1450 1460 1470 1480 1490  
1500 1510 1520 1530 1540 1550 1560 1570 1580 1590  
1600 1610 1620 1630 1640 1650 1660 1670 1680 1690  
1700 1710 1720 1730 1740 1750 1760 1770 1780 1790  
1800 1810 1820 1830 1840 1850 1860 1870 1880 1890  
1900 1910 1920 1930 1940 1950 1960 1970 1980 1990  
2000 2010 2020 2030 2040 2050 2060 2070 2080 2090  
2100 2110 2120 2130 2140 2150 2160 2170 2180 2190  
2200 2210 2220 2230 2240 2250 2260 2270 2280 2290  
2300 2310 2320 2330 2340 2350 2360 2370 2380 2390  
2400 2410 2420 2430 2440 2450 2460 2470 2480 2490  
2500 2510 2520 2530 2540 2550 2560 2570 2580 2590  
2600 2610 2620 2630 2640 2650 2660 2670 2680 2690  
2700 2710 2720 2730 2740 2750 2760 2770 2780 2790  
2800 2810 2820 2830 2840 2850 2860 2870 2880 2890  
2900 2910 2920 2930 2940 2950 2960 2970 2980 2990  
3000 3010 3020 3030 3040 3050 3060 3070 3080 3090  
3100 3110 3120 3130 3140 3150 3160 3170 3180 3190  
3200 3210 3220 3230 3240 3250 3260 3270 3280 3290  
3300 3310 3320 3330 3340 3350 3360 3370 3380 3390  
3400 3410 3420 3430 3440 3450 3460 3470 3480 3490  
3500 3510 3520 3530 3540 3550 3560 3570 3580 3590  
3600 3610 3620 3630 3640 3650 3660 3670 3680 3690  
3700 3710 3720 3730 3740 3750 3760 3770 3780 3790  
3800 3810 3820 3830 3840 3850 3860 3870 3880 3890  
3900 3910 3920 3930 3940 3950 3960 3970 3980 3990  
4000 4010 4020 4030 4040 4050 4060 4070 4080 4090  
4100 4110 4120 4130 4140 4150 4160 4170 4180 4190  
4200 4210 4220 4230 4240 4250 4260 4270 4280 4290  
4300 4310 4320 4330 4340 4350 4360 4370 4380 4390  
4400 4410 4420 4430 4440 4450 4460 4470 4480 4490  
4500 4510 4520 4530 4540 4550 4560 4570 4580 4590  
4600 4610 4620 4630 4640 4650 4660 4670 4680 4690  
4700 4710 4720 4730 4740 4750 4760 4770 4780 4790  
4800 4810 4820 4830 4840 4850 4860 4870 4880 4890  
4900 4910 4920 4930 4940 4950 4960 4970 4980 4990  
5000 5010 5020 5030 5040 5050 5060 5070 5080 5090  
5100 5110 5120 5130 5140 5150 5160 5170 5180 5190  
5200 5210 5220 5230 5240 5250 5260 5270 5280 5290  
5300 5310 5320 5330 5340 5350 5360 5370 5380 5390  
5400 5410 5420 5430 5440 5450 5460 5470 5480 5490  
5500 5510 5520 5530 5540 5550 5560 5570 5580 5590  
5600 5610 5620 5630 5640 5650 5660 5670 5680 5690  
5700 5710 5720 5730 5740 5750 5760 5770 5780 5790  
5800 5810 5820 5830 5840 5850 5860 5870 5880 5890  
5900 5910 5920 5930 5940 5950 5960 5970 5980 5990  
6000 6010 6020 6030 6040 6050 6060 6070 6080 6090  
6100 6110 6120 6130 6140 6150 6160 6170 6180 6190  
6200 6210 6220 6230 6240 6250 6260 6270 6280 6290  
6300 6310 6320 6330 6340 6350 6360 6370 6380 6390  
6400 6410 6420 6430 6440 6450 6460 6470 6480 6490  
6500 6510 6520 6530 6540 6550 6560 6570 6580 6590  
6600 6610 6620 6630 6640 6650 6660 6670 6680 6690  
6700 6710 6720 6730 6740 6750 6760 6770 6780 6790  
6800 6810 6820 6830 6840 6850 6860 6870 6880 6890  
6900 6910 6920 6930 6940 6950 6960 6970 6980 6990  
7000 7010 7020 7030 7040 7050 7060 7070 7080 7090  
7100 7110 7120 7130 7140 7150 7160 7170 7180 7190  
7200 7210 7220 7230 7240 7250 7260 7270 7280 7290  
7300 7310 7320 7330 7340 7350 7360 7370 7380 7390  
7400 7410 7420 7430 7440 7450 7460 7470 7480 7490  
7500 7510 7520 7530 7540 7550 7560 7570 7580 7590  
7600 7610 7620 7630 7640 7650 7660 7670 7680 7690  
7700 7710 7720 7730 7740 7750 7760 7770 7780 7790  
7800 7810 7820 7830 7840 7850 7860 7870 7880 7890  
7900 7910 7920 7930 7940 7950 7960 7970 7980 7990  
8000 8010 8020 8030 8040 8050 8060 8070 8080 8090  
8100 8110 8120 8130 8140 8150 8160 8170 8180 8190  
8200 8210 8220 8230 8240 8250 8260 8270 8280 8290  
8300 8310 8320 8330 8340 8350 8360 8370 8380 8390  
8400 8410 8420 8430 8440 8450 8460 8470 8480 8490  
8500 8510 8520 8530 8540 8550 8560 8570 8580 8590  
8600 8610 8620 8630 8640 8650 8660 8670 8680 8690  
8700 8710 8720 8730 8740 8750 8760 8770 8780 8790  
8800 8810 8820 8830 8840 8850 8860 8870 8880 8890  
8900 8910 8920 8930 8940 8950 8960 8970 8980 8990  
9000 9010 9020 9030 9040 9050 9060 9070 9080 9090  
9100 9110 9120 9130 9140 9150 9160 9170 9180 9190  
9200 9210 9220 9230 9240 9250 9260 9270 9280 9290  
9300 9310 9320 9330 9340 9350 9360 9370 9380 9390  
9400 9410 9420 9430 9440 9450 9460 9470 9480 9490  
9500 9510 9520 9530 9540 9550 9560 9570 9580 9590  
9600 9610 9620 9630 9640 9650 9660 9670 9680 9690  
9700 9710 9720 9730 9740 9750 9760 9770 9780 9790  
9800 9810 9820 9830 9840 9850 9860 9870 9880 9890  
9900 9910 9920 9930 9940 9950 9960 9970 9980 9990  
10000 10010 10020 10030 10040 10050 10060 10070 10080 10090  
10100 10110 10120 10130 10140 10150 10160 10170 10180 10190  
10200 10210 10220 10230 10240 10250 10260 10270 10280 10290  
10300 10310 10320 10330 10340 10350 10360 10370 10380 10390  
10400 10410 10420 10430 10440 10450 10460 10470 10480 10490  
10500 10510 10520 10530 10540 10550 10560 10570 10580 10590  
10600 10610 10620 10630 10640 10650 10660 10670 10680 10690  
10700 10710 10720 10730 10740 10750 10760 10770 10780 10790  
10800 10810 10820 10830 10840 10850 10860 10870 10880 10890  
10900 10910 10920 10930 10940 10950 10960 10970 10980 10990  
11000 11010 11020 11030 11040 11050 11060 11070 11080 11090  
11100 11110 11120 11130 11140 11150 11160 11170 11180 11190  
11200 11210 11220 11230 11240 11250 11260 11270 11280 11290  
11300 11310 11320 11330 11340 11350 11360 11370 11380 11390  
11400 11410 11420 11430 11440 11450 11460 11470 11480 11490  
11500 11510 11520 11530 11540 11550 11560 11570 11580 11590  
11600 11610 11620 11630 11640 11650 11660 11670 11680 11690  
11700 11710 11720 11730 11740 11750 11760 11770 11780 11790  
11800 11810 11820 11830 11840 11850 11860 11870 11880 11890  
11900 11910 11920 11930 11940 11950 11960 11970 11980 11990  
12000 12010 12020 12030 12040 12050 12060 12070 12080 12090  
12100 12110 12120 12130 12140 12150 12160 12170 12180 12190  
12200 12210 12220 12230 12240 12250 12260 12270 12280 12290  
12300 12310 12320 12330 12340 12350 12360 12370 12380 12390  
12400 12410 12420 12430 12440 12450 12460 12470 12480 12490  
12500 12510 12520 12530 12540 12550 12560 12570 12580 12590  
12600 12610 12620 12630 12640 12650 12660 12670 12680 12690  
12700 12710 12720 12730 12740 12750 12760 12770 12780 12790  
12800 12810 12820 12830 12840 12850 12860 12870 12880 12890  
12900 12910 12920 12930 12940 12950 12960 12970 12980 12990  
13000 13010 13020 13030 13040 13050 13060 13070 13080 13090  
13100 13110 13120 13130 13140 13150 13160 13170 13180 13190  
13200 13210 13220 13230 13240 13250 13260 13270 13280 13290  
13300 13310 13320 13330 13340 13350 13360 13370 13380 13390  
13400 13410 13420 13430 13440 13450 13460 13470 13480 13490  
13500 13510 13520 13530 13540 13550 13560 13570 13580 13590  
13600 13610 13620 13630 13640 13650 13660 13670 13680 13690  
13700 13710 13720 13730 13740 13750 13760 13770 13780 13790  
13800 13810 13820 13830 13840 13850 13860 13870 13880 13890  
13900 13910 13920 13930 13940 13950 13960 13970 13980 13990  
14000 14010 14020 14030 14040 14050 14060 14070 14080 14090  
14100 14110 14120 14130 14140 14150 14160 14170 14180 14190  
14200 14210 14220 14230 14240 14250 14260 14270 14280 14290  
14300 14310 14320 14330 14340 14350 14360 14370 14380 14390  
14400 14410 14420 14430 14440 14450 14460 14470 14480 14490  
14500 14510 14520 14530 14540 14550 14560 14570 14580 14590  
14600 14610 14620 14630 14640 14650 14660 14670 14680 14690  
14700 14710 14720 14730 14740 14750 14760 14770 14780 14790  
14800 14810 14820 14830 14840 14850 14860 14870 14880 14890  
14900 14910 14920 14930 14940 14950 14960 14970 14980 14990  
15000 15010 15020 15030 15040 15050 15060 15070 15080 15090  
15100 15110 15120 15130 15140 15150 15160 15170 15180 15190  
15200 15210 15220 15230 15240 15250 15260 15270 15280 15290  
15300 15310 15320 15330 15340 15350 15360 15370 15380 15390  
15400 15410 15420 15430 15440 15450 15460 15470 15480 15490  
15500 15510 15520 15530 15540 15550 15560 15570 15580 15590  
15600 15610 15620 15630 15640 15650 15660 15670 15680 15690  
15700 15710 15720 15730 15740 15750 15760 15770 15780 15790  
15800 15810 15820 15830 15840 15850 15860 15870 15880 15890  
15900 15910 15920 15930 15940 15950 15960 15970 15980 15990  
16000 16010 16020 16030 16040 16050 16060 16070 16080 16090  
16100 16110 16120 16130 16140 16150 16160 16170 16180 16190  
16200 16210 16220 16230 16240 16250 16260 16270 16280 16290  
16300 16310 16320 16330 16340 16350 16360 16370 16380 16390  
16400 16410 16420 16430 16440 16450 16460 16470 16480 16490  
16500 16510 16520 16530 16540 16550 16560 16570 16580 16590  
16600 16610 16620 16630 16640 16650 16660 16670 16680 16690  
16700 16710 16720 16730 16740 16750 16760 16770 16780 16790  
16800 16810 16820 16830 16840 16850 16860 16870 16880 16890  
16900 16910 16920 16930 16940 16950 16960 16970 16980 16990  
17000 17010 17020 17030 17040 17050 17060 17070 17080 17090  
17100 17110 17120 17130 17140 17150 17160 17170 17180 17190  
17200 17210 17220 17230 17240 17250 17260 17270 17280 17290  
17300 17310 17320 17330 17340 17350 17360 17370 17380 17390  
17400 17410 17420 17430 17440 17450 17460 17470 17480 17490  
17500 17510 17520 17530 17540 17550 17560 17570 17580 17590  
17600 17610 17620 17630 17640 17650 17660 17670 17680 17690  
17700 17710 17720 17730 17740 17750 17760 17770 17780 17790  
17800 17810 17820 17830 17840 17850 17860 17870 17880 17890  
17900 17910 17920 17930 17940 17950 17960 17970 17980 17990  
18000 18010 18020 18030 18040 18050 18060 18070 18080 18090  
18100 18110 18120 18130 18140 18150 18160 18170 18180 18190  
18200 18210 18220 18230 18240 18250 18260 18270 18280 18290  
18300 18310 18320 18330 18340 18350 18360 18370 18380 18390  
18400 18410 18420 18430 18440 18450 18460 18470 18480 18490  
18500 18510 18520 18530 18540 18550 18560 18570 18580 18590  
18600 18610 18620 18630 18640 18650 18660 18670 18680 18690  
18700 18710 18720 18730 18740 18750 18760 18770 18780 18790  
18800 18810 18820 18830 18840 18850 18860 18870 18880 18890  
18900 18910 18920 18930 18940 18950 18960 18970 18980 18990  
19000 19010 19020 19030 19040 19050 19060 19070 19080 19090  
19100 19110 19120 19130 19140 19150 19

60 70 80 90 100 120  
ACGAGGAATCCACAAGGAGACGCGCTCGAGGCTGGGTATCAGCATCGAAAAGGGCTTTG  
||| ||| ||| ||| ||| |||  
GCGGAATCTCTCDCTNCCDGCVARRTCAC  
X 10 20 30

3. US-09-235-416-2 (1327-1803)  
US-09-235-416-6 Sequence 6, Application US/09235416A

Initial Score = 8 Optimized Score = 15 Significance = -0.40  
Residue Identity = 33% Matches = 15 Mismatches = 15  
Gaps = 0 Conservative Substitutions = 0

80 90 100 110 120 130 140  
AACGAGAAGCGCGCTCGAGGAGTGGCTATCAGCATCGAAAGGGCTTTGTCGCCCTTACCACATCAA  
||| ||| ||| ||| ||| |||  
GCGGAATCTCTCDGANCCDGCVARRTCAC  
X 10 20 30

4. US-09-235-416-2 (1327-1803)  
US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score = 6 Optimized Score = 9 Significance = -0.41  
Residue Identity = 42% Matches = 9 Mismatches = 12  
Gaps = 0 Conservative Substitutions = 0

40 50 60 80 90  
CAGACCTGGGAGAGAGCTGCCAAGACCGAGGAATCCACAGGACGAGAGCGCGC  
||| ||| ||| ||| |||  
ATGTCGGCGGTGGAATATC  
X 10 20

5. US-09-235-416-2 (1327-1803)  
US-09-235-416-4 Sequence 4, Application US/09235416A

Initial Score = 5 Optimized Score = 12 Significance = -0.42  
Residue Identity = 52% Matches = 12 Mismatches = 11  
Gaps = 0 Conservative Substitutions = 0

250 260 X 270 280 X 290 300  
TCACCAAGATACACAAGCGGAATTCGTCTGACGCTTCGAGATCCTGAAGACACATGTA  
||| ||| ||| ||| |||  
GAATTCCTGCTCGCTGTTTCA  
X 10 20 X





Residue Identity = 36% Matches = 15 Mismatches = 15  
Gaps = 0 Conservative Substitutions = 0

250 260 270 280 290 X 310  
TATTTGACGATGTTTCAGAAAGCGCGGCGCTGCTGGCTGGTCGAAGACAACGAAGATAGCGATTTC  
||||| | | | | | | | |  
GGCGGAATTCCTCDCTNCCDCSVARRTCNAC  
X 10 20 30

## 3. US-09-235-416-2 (1804-2352)

US-09-235-416-4 Sequence 4, Application US/09235416A

Initial Score = 9 Optimized Score = 11 Significance = -0.40  
Residue Identity = 47% Matches = 11 Mismatches = 12  
Gaps = 0 Conservative Substitutions = 0

410 420 X 430 440 X 450 460  
GGGTACCCCTCGTAGTGATGACGACGCGGTGAGCGCGCTGTTTTTTGGTGATAAGAAGTCGAAACA  
|| |||||  
GAATTCCTGCTTCGCTGTTTTCA  
X 10 20 X

## 4. US-09-235-416-2 (1804-2352)

US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score = 7 Optimized Score = 9 Significance = -0.41  
Residue Identity = 42% Matches = 9 Mismatches = 12  
Gaps = 0 Conservative Substitutions = 0

120 130 140 150 X 170  
TCAGATTCTCCTTTGCCGCGACTTTCGTGGAAGGATAGCGACTGGTTCTATGCTTCGAGGG  
| | | | |  
ATGTCGGGCGGTGGAATATC  
X 10 20







Residue Identity = 43% Matches = 17 Mismatches = 13  
Gaps = 0 Conservative Substitutions = 0

210 220 230 240 250 270  
GCTCCAGTAGCTCCGGTGGAGTGTCTCGTCAGAACCCGCCAAATCTACGACACTGATCTTCGCAACCT  
||||| || ||||| |||  
GCGGGAATCTCDDCTNCCDGCVARRTCAC  
X 10 20 30

3. US-09-235-416-2' (1327-1803)  
US-09-235-416-6 Sequence 6, Application US/09235416A

Initial Score = 12 Optimized Score = 19 Significance = -0.20  
Residue Identity = 50% Matches = 19 Mismatches = 11  
Gaps = 0 Conservative Substitutions = 0

210 220 230 240 250 270  
GCTCCAGTAGCTCCGGTGGAGTGTCTCGTCAGAACCCGCCAAATCTACGACACTGATCTTCGCAACCT  
||||| || ||||| |||  
GCGGGAATCTCDDGANCDCGCVARRTCAC  
X 10 20 30

4. US-09-235-416-2' (1327-1803)  
US-09-235-416-5 Sequence 5, Application US/09235416A

Initial Score = 9 Optimized Score = 13 Significance = -0.49  
Residue Identity = 30% Matches = 13 Mismatches = 17  
Gaps = 0 Conservative Substitutions = 0

240 250 260 270 280 290 300  
CGCTCAGAACCCGCCAAATCTACGACACTGATCTTCGCAACCTTCTCTGTGCCATTTGGTCTCTTCAT  
|| ||||| |||  
GCGCGGATCCATYTYGCHTAYGGNCARAC  
X 10 20 30

5. US-09-235-416-2' (1327-1803)  
US-09-235-416-4 Sequence 4, Application US/09235416A

Initial Score = 8 Optimized Score = 12 Significance = -0.59  
Residue Identity = 52% Matches = 12 Mismatches = 11  
Gaps = 0 Conservative Substitutions = 0

250 260 270 280 290 300 310  
GCCAAATCTACGACACTGATCTTCGCAACCTTCTCTGTGCCATTTGGTCTCTTCATCATGC  
|| ||||| |||  
GAATTCCTGCTTCGCTGTTTCA  
X 10 20 X

6. US-09-235-416-2' (1327-1803)  
US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score = 7 Optimized Score = 10 Significance = -0.69  
Residue Identity = 47% Matches = 10 Mismatches = 11  
Gaps = 0 Conservative Substitutions = 0

200 210 220 230 240 250  
CCCTCCTTCAGTCCGGCTCCAGTAGTCCGGTGGAGCTTGTCTCGCTCAGAACCCGCCAAAT  
|| ||||| |||  
ATGTCGGCGGTGGAAATATC  
X 10 20







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 04:39:12 ; Search time 2969.55 Seconds  
(without alignments)  
104.538 Million cell updates/sec

Title: US-09-235-416-4  
Perfect score: 23  
Sequence: 1 gaattcgtcttcgtgtttca 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	81.7	396	12	AZ390315
2	18.8	81.7	747	12	AZ804429
3	18.8	81.7	882	12	AZ685448
4	18.2	79.1	178	9	AA253238
5	18.2	79.1	224	9	AW844301
6	18.2	79.1	280	10	T06084
7	18.2	79.1	281	9	AA378449
8	18.2	79.1	293	9	AW844294
9	18.2	79.1	303	9	AA355871
10	18.2	79.1	306	9	AA411640
11	18.2	79.1	322	10	W25620
12	18.2	79.1	328	10	BF845084
13	18.2	79.1	337	12	AQ260557
14	18.2	79.1	356	10	BI033412
15	18.2	79.1	382	9	AV721381
16	18.2	79.1	411	10	BF995757
17	18.2	79.1	413	12	AQ333260

C 18	18.2	79.1	446	10	W89640
19	18.2	79.1	447	9	AW844305
20	18.2	79.1	459	10	R56708
21	18.2	79.1	465	10	BF028620
22	18.2	79.1	468	9	AA773267
23	18.2	79.1	468	10	H94360
C 24	18.2	79.1	473	10	BE764542
25	18.2	79.1	547	10	W05924
C 26	18.2	79.1	567	10	BG082824
27	18.2	79.1	611	9	AI730423
C 28	18.2	79.1	621	10	BM311296
29	18.2	79.1	623	10	BE540124
C 30	18.2	79.1	638	12	BH033255
C 31	18.2	79.1	644	9	AW304933
32	18.2	79.1	660	9	AI726774
33	18.2	79.1	663	10	BF115925
34	18.2	79.1	669	10	BG926285
C 35	18.2	79.1	705	9	AW131625
36	18.2	79.1	724	10	BG329673
C 37	18.2	79.1	729	10	BF096178
C 38	18.2	79.1	743	9	AA027726
C 39	18.2	79.1	756	9	AA206424
40	18.2	79.1	775	10	BI858782
41	18.2	79.1	780	10	BI335414
42	18.2	79.1	812	9	AJ396274
C 43	18.2	79.1	855	9	AL552868
C 44	18.2	79.1	922	10	BE612931
C 45	18.2	79.1	972	9	AL571242

ALIGNMENTS

RESULT 1  
AZ390315  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ390315 396 bp DNA linear GSS 03-OCT-2000  
IM0151C07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0151C07 R, DNA sequence.  
AZ390315  
AZ390315.1 GI:10505054  
GSS.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 396)  
Dunn, C., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0151 row: C column: 07  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 396.  
Location/Qualifiers  
1. 396  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0151C07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"

```

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid p1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      54 a 75 c 85 g 182 t
ORIGIN

```

```

Query Match      81.7%; Score 18.8; DB 12; Length 396;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattcgtctgcgtgttttc 22
|||||
Db 65 GAATTCCTGCTCTCTGCTTC 86

```

## RESULT 2

```

A2804429
LOCUS      747 bp      DNA      linear      GSS 16-FEB-2001
DEFINITION 2M0065H14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0065H14 F, DNA sequence.
ACCESSION  A2804429
VERSION     A2804429.1 GI:12956752
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 747)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0065 row: H column: 14
Seq primer: CTTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 747.
Location/Qualifiers
1..747
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0065H14"

```

## FEATURES

source

```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid p1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      167 a 154 c 188 g 238 t
ORIGIN

Query Match      81.7%; Score 18.8; DB 12; Length 747;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aattcgtctgcgtgttttc 23
|||||
Db 573 ATTTCCTGCTCTGCTGTAA 594

RESULT 3
A2685448
LOCUS      882 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION ENTJY17TF Entamoeba histolytica sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION  A2685448
VERSION     A2685448.1 GI:11822594
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica
REFERENCE   1 (bases 1 to 882)
AUTHORS     Loftus,B., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL     Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@futsig.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 3
High quality sequence stop: 766.
Location/Qualifiers
1..882
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHD1; Site:1; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica"

```

## FEATURES

source

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."

BASE COUNT 298 a 153 c 126 g 305 t  
ORIGIN

Query Match 81.7%; Score 18.8; DB 12; Length 882;  
Best Local Similarity 90.9%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aattctgctgcgtgttttca 23  
||||| ||||| ||||| |||||

Db 722 AATCTGTCTCACTGTTTCA 743

RESULT 4

AA253238

LOCUS AA253238 178 bp mRNA linear EST 06-AUG-1997  
DEFINITION zr53f07.r1 Soares\_NhMpu\_S1 Homo sapiens cDNA clone IMAGE:667141 5' similar to TR:G1063396 G1063396 RAGA. ;, mRNA sequence.

ACCESSION AA253238

VERSION AA253238.1 GI:1885404

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 178)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-Merck EST Project 1997

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.lml.gov) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Insert Length: 598 Std Error: 0.00

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1. .178

/organism="Homo sapiens"

/db\_xref="GDB:5561108"

/db\_xref="taxon:9606"

/clone="IMAGE:667141"

/clone\_lib="Soares\_NhMpu\_S1"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

/note="Organ: mixed (see below); Vector: pT73D-Pac

(Pharmacia) with a modified polylinker; Site\_1: Not I;

Site\_2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NDHM, pregnant uterus

NBHPU, and fetal heart NBHH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 42 a 50 c 40 g 46 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 178;  
Best Local Similarity 87.0%; Pred. No. 6.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtctgcgtgttttca 23  
||||| ||||| ||||| |||||

Db 4 GAATTCGAACCTGCTGCTTTCA 26

RESULT 5

AW844301

LOCUS AW844301 224 bp mRNA linear EST 18-MAY-2000  
DEFINITION RC4-CN0050-130200-012-f01 CN0050 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW844301

VERSION AW844301.1 GI:7938284

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 224)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC4-CN0050-130

200-012-f01&t3=2000-02-13&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 224.

Location/Qualifiers

1. .224

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CN0050"

/dev\_stage="Adult"

/note="Organ: colon normal; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

fragments into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

62 a 59 c 50 g 53 t

BASE COUNT

ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 224;

Best Local Similarity 87.0%; Pred. No. 7e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23  
 ||||||| ||||||| |||||||  
 Db 144 GAATTCCAACTCGCTGCTTTCA 166

RESULT 6  
 T06084 280 bp mRNA linear EST 30-JUN-1993  
 DEFINITION clone HFBDO22, mRNA sequence.

ACCESSION T06084  
 VERSION T06084.1 GI:317233  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 280)  
 AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.  
 TITLE 3,400 expressed sequence tags identify diversity of transcripts  
 from human brain

JOURNAL Nature Genet. 4, 256-267 (1993)  
 MEDLINE 93364420  
 COMMENT Contact: Adams, MD  
 The Institute for Genomic Research  
 932 Clopper Road, Gaithersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: mcdams@tigr.org  
 Seq primer: M13-21.

FEATURES  
 source Location/Qualifiers

1..280  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):82755"  
 /db\_xref="taxon:9606"  
 /clone="HFBDO22"  
 /clone\_lib="Fetal brain, Stratagene (cat#936206)"  
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
 oligo-qt + random primed cDNA synthesis; lambdaZAP-II  
 vector, 1.0kb average inser size."

BASE COUNT 66 a 64 c 80 t 10 others  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 280;  
 Best Local Similarity 87.0%; Pred. No. 7.2e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23  
 ||||||| ||| ||||||| ||  
 Db 81 GAATTCCTCCTGGCTGTTTCCA 103

RESULT 7  
 AA378449 281 bp mRNA linear EST 21-APR-1997  
 LOCUS EST91145 Synovial sarcoma Homo sapiens cDNA 5' end similar to  
 DEFINITION guanine nucleotide-binding protein raga, mRNA sequence.

ACCESSION AA378449  
 VERSION AA378449.1 GI:2030768  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 281)  
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult  
 ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White  
 ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
 Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald  
 ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,  
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,

Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
 Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,  
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,  
 Kunsch,C., Hungjun,J., Li,H., Melssner,P.S., Olsen,H., Raymond,L.,  
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon  
 ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and  
 Venter,J.C.

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)

96026280  
 COMMENT Other ESTs: THC189208  
 Contact: Kerlavage, AR  
 Bioinformatics

The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423

Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 source Location/Qualifiers

1..281  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):182850"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Synovial sarcoma"  
 /sex="female"  
 /tissue\_type="synovial membrane"  
 /dev\_stage="adult, 20 yrs"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 77 a 72 c 61 g 71 t  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 281;  
 Best Local Similarity 87.0%; Pred. No. 7.2e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23  
 ||||||| ||||||| |||||||  
 Db 229 GAATTCCAACTCGCTGCTTCA 251

RESULT 8  
 AW844294 293 bp mRNA linear EST 18-MAY-2000  
 LOCUS RC4-CN0050-130200-012-c08 CN0050 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION  
 ACCESSION AW844294  
 VERSION AW844294.1 GI:7938277  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 293)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663

TITLE  
 JOURNAL  
 MEDLINE

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC4-CN0050-130  
200-012-c086t3-2000-02-136t4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 9  
High quality sequence stop: 293.

FEATURES source  
1. .293  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CN0050"  
/dev\_stage="Adult"  
/note="Organ: colon\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
76 a 81 c 64 g 72 t

Query Match 79.1%; Score 18.2; DB 9; Length 293;  
Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgctgcgtgttttca 23  
||||| ||||||| ||||||| |||||||  
Db 158 GAATTCCAACTTCGCTGCTTCA 180

RESULT 9  
AA355871  
LOCUS AA355871 303 bp mRNA linear EST 21-APR-1997  
DEFINITION EST64335 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to  
guanine nucleotide-binding protein raga, mRNA sequence.  
AA355871  
VERSION AA355871.1 GI:2008190  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 303)  
ADAMS, M.D., Kerkness, E.F., Weinstock, K.G., Gocayne, J.D., White  
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White  
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,  
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald  
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,  
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palanco, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,  
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,  
Wei, F.F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon  
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and  
Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)  
MEDLINE 96026280  
COMMENT Other\_ESTs: THC189208  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (http://www.tigr.org/tldb/hngi/hngi.html)  
Seq primer: M13 Reverse.

FEATURES source  
1. .303  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):157999"  
/db\_xref="taxon:9606"  
/clone\_lib="Jurkat T-cells VI"  
/cell\_type="T-lymphocyte"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 73 a 75 c 60 g 93 t 2 others  
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 303;  
Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgctgcgtgttttca 23  
||||| ||||||| ||||||| |||||||  
Db 54 GAATTCCAACTTCGCTGCTTCA 76

RESULT 10  
AA411640  
LOCUS AA411640 306 bp mRNA linear EST 17-MAY-1997  
DEFINITION zvl5c03.r1 Soares\_NhMpu\_S1 Homo sapiens cDNA clone IMAGE:753700 5'  
similar to TR:G1063396 G1063396 RAGA. ;, mRNA sequence.  
AA411640  
VERSION AA411640.1 GI:2069365  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 306)  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie  
T., Waterston, R. and Willson, R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK

TITLE Washington University School of Medicine  
JOURNAL 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
COMMENT Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 305.

FEATURES source  
1. .306  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:597656"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares\_NhMpu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT 81 a 76 c 71 g 78 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 306;  
Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtgtttca 23  
||||||| ||||||| |||||  
Db 102 GAATTCCAACTTCGCTGCTTCA 124

RESULT 11  
LOCUS W25620 322 bp mRNA linear EST 20-AUG-1996  
DEFINITION z67le02.rl Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
IMAGE:309050 5' similar to WF:T24Fl.1 CE02361 GTP-BINDING PROTEIN  
;; mRNA sequence.

ACCESSION W25620  
VERSION W25620.1 GI:1303455  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 322)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins  
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore  
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,  
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisakis,E.,  
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 738 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 234.

FEATURES  
source  
1..322  
/organism="Homo sapiens"  
/db\_xref="GDB:1252463"  
/db\_xref="taxon:9606"  
/clone="IMAGE:309050"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTACCACTCAATCAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],

double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBH19W."

BASE COUNT 86 a 84 c 70 g 77 t 5 others  
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 322;  
Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtgtttca 23  
||||||| ||||||| |||||  
Db 145 GAATTCCAACTTCGCTGCTTCA 167

RESULT 12  
LOCUS BF845084/c 328 bp mRNA linear EST 13-JAN-2001  
DEFINITION CM1-HT1145-271200-677-d08 HT1145 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF845084  
VERSION BF845084.1 GI:12201298  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 328)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1st2-CM1-HT1145-  
271200-677-d08&t3=2000-12-27&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 8  
High quality sequence stop: 327.  
Location/Qualifiers

FEATURES  
source  
1..328  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT1145"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
79 a 75 c 85 t  
BASE COUNT 79 a 75 c 85 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 328;  
 Best Local Similarity 87.0%; Pred. No. 7.4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtcttcgtgttttca 23  
 ||||| ||||| ||||| ||||| |||||  
 Db 208 GAATTCCAACTTCGCTGCTTCA 186

RESULT 13  
 LOCUS AQ260557 337 bp DNA linear GSS 24-OCT-1998  
 DEFINITION CITBI-El-2508E24.TF CITBI-El Homo sapiens genomic clone 2508E24,  
 DNA sequence.  
 ACCESSION AQ260557  
 VERSION AQ260557.1 GI:3787081  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 337)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
 Venter,J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready

JOURNAL Map Building  
 COMMENT Unpublished (1998)  
 Other\_GSSs: CITBI-El-2508E24.TR  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13-21  
 Class: BAC ends.

FEATURES  
 source  
 1..337  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2508E24"  
 /clone\_lib="CITBI-El"  
 /sex="male"  
 /cell\_type="sperm"  
 /note="Vector: pBelobAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Caltech Human BAC Library D"

BASE COUNT 98 a 67 c 75 g 97 t  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 12; Length 337;  
 Best Local Similarity 87.0%; Pred. No. 7.4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtcttcgtgttttca 23  
 ||||| ||||| ||||| ||||| |||||  
 Db 52 GAATTCCTCTCTGCTGCTTTTCA 74

RESULT 14  
 BI033412/c  
 LOCUS PM1-NN1207-150201-023-b07 NN1207 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION  
 ACCESSION BI033412  
 VERSION BI033412.1 GI:14440038  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 356)  
 AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-NN1207-  
 150201-023-b07&t3=2001-02-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 16  
 High quality sequence stop: 317.

FEATURES  
 source  
 1..356  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NN1207"  
 /dev\_stage="Adult"  
 /note="Organ: nervous\_normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 54 a 77 c 92 g 93 t  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 356;  
 Best Local Similarity 87.0%; Pred. No. 7.4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtcttcgtgttttca 23  
 ||||| ||||| ||||| ||||| |||||  
 Db 93 GAATTCCAACTTCGCTGCTTCA 71

RESULT 15  
 AV721381  
 LOCUS AV721381 HTB Homo sapiens cDNA clone HTBADE08 5', mRNA sequence.  
 DEFINITION  
 ACCESSION AV721381  
 VERSION AV721381.1 GI:10818533  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 382)  
 AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,  
 Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu  
 ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,  
 Chen,J., Chen,Z. and Han,Z.  
 TITLE Homo sapiens cDNA HTB clones

JOURNAL Unpublished (2000)  
COMMENT Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@hgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
1. .382  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HTBADE08"  
/clone\_lib="HTB"  
/tissue\_type="Hypothalamus"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pbluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 81 a 99 c 78 g 124 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 382;  
Best Local Similarity 87.0%; Pred. No. 7.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gaattcctgcttcgtgtttca 23  
||||||| ||||||| |||||  
Db 70 GAATTCACACTCGCTGCTTCA 92

Search completed: July 25, 2002, 04:39:19  
Job time: 3157 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 04:01:22 ; Search time 63.5 Seconds  
(without alignments)  
1371.368 Million cell updates/sec

Title: US-09-235-416-1

Perfect score: 4030

Sequence: 1 MSGGNIKVVRRFPNARE.....ELRQQQAQMEALKTAQGEF 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4030	100.0	784	20	Thermomyces lanugi
2	1688.5	41.4	1816	21	Human kinesin-like
3	1660.5	41.2	893	22	Human polypeptide
4	1658	41.1	1103	21	Human KLIMP protei
5	1658	41.1	1103	22	Human kinesin-like
6	1635	40.6	1773	22	Drosophila melanog
7	1430	35.5	757	22	Human diagnostic a
8	1396.5	34.7	1921	22	Drosophila melanog
9	1278.5	31.7	504	21	Gene 5 human secre
10	1276	31.7	503	21	Human secreted pro
11	1251	31.0	421	22	Human polypeptide

12	1128	28.0	1174	22	AB61704	Drosophila melanog
13	1035	25.7	1121	22	AB672021	Drosophila melanog
14	814	20.2	955	15	AAR57365	K39 polypeptide of
15	814	20.2	955	17	AAW03691	Leishmania chagasi
16	781.5	19.4	677	22	AB65183	Drosophila melanog
17	750.5	17.2	784	22	AB671112	Drosophila melanog
18	691.5	17.2	963	22	AAW78880	Human protein seq
19	683.5	17.0	979	22	AAW79864	Human protein seq
20	679	16.8	1212	22	AB671114	Drosophila melanog
21	673	16.7	1057	22	AAG67419	Amino acid sequenc
22	673	16.7	1057	22	AAB47212	Human KSP. Homo s
23	671.5	16.7	513	22	AAB47214	Human KSP-K491. S
24	671.5	16.7	575	22	AAB47215	Human KSP-S553. S
25	668	16.6	2633	22	ABG06505	Novel human diagno
26	663	16.5	2663	22	AAW39097	Human polypeptide
27	662.5	16.4	2954	20	AAW01632	Amino acid sequenc
28	651.5	16.2	975	22	AB63485	Drosophila melanog
29	649.5	16.1	975	19	AAW72746	Drosophila kinesin
30	645.5	16.0	2688	22	AAW40883	Human polypeptide
31	640	15.9	1003	22	AB61405	Drosophila melanog
32	633	15.7	1184	22	AAG67415	Amino acid sequenc
33	630	15.6	814	22	AB65317	Drosophila melanog
34	628	15.6	368	22	AAB47216	Human KSP-K368. S
35	625	15.5	411	19	AAW72745	Drosophila kinesin
36	625	15.5	441	19	AAW72744	Drosophila kinesin
37	616.5	15.3	382	22	AAB47213	Human KSP-L360. Sy
38	610.5	15.1	1066	22	AAG67418	Amino acid sequenc
39	603.5	15.0	1048	22	AB659245	Drosophila melanog
40	601	14.9	1029	22	ABG20649	Novel human diagno
41	587.5	14.6	1038	22	AAG67416	Amino acid sequenc
42	584	14.5	1518	21	AAG40075	Arabidopsis thalia
43	584	14.5	1662	21	AAG31282	Arabidopsis thalia
44	580.5	14.4	1460	21	AAG40077	Arabidopsis thalia
45	580.5	14.4	1462	21	AAG40076	Arabidopsis thalia

## ALIGNMENTS

RESULT 1

AAW06618

ID AAY06618 standard; Protein; 784 AA.

XX

AC AAY06618;

XX

DT 26-OCT-1999 (first entry)

XX

DE Thermomyces lanuginosus kinesin motor protein TL-gamma.

XX

KW TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;

KW neurodegenerative disease; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis.

XX

OS Thermomyces lanuginosus.

XX

PN WO9937659-A1.

XX

PD 29-JUL-1999.

XX

PF 22-JAN-1999; 99WO-US01355.

XX

PR 23-JAN-1998; 98US-0072361.

XX

PA (REGC ) UNIV CALIFORNIA.

XX

PI Goldstein LSB, Sakowicz R;

XX

DR WPI; 1999-493950/41.

DR N-PSDB; AAX87656.

XX

PT New nucleic acid encoding microtubule motor protein, used for diagnosis of fungal infection and neurodegenerative disease



QY 161 HPSTGPPYEDLAKLVVRVSRFQETENLMDGNGKARTVAATNMNETSSRSRSHAVFTLTQKWH 240  
DB 171 hpllgpyvedisklavtydiadlmdagnkartvaatnmnetsrshavftltqkhh 230  
QY 241 DEETKMDTEKAKISLVLDLAGSERATSGATGARLKEGAENRSLSTLGRVIAALADM-- 298  
DB 231 dnetnlstekvskislvdlagseradstgagtrikeganlnkslttlgkvisalaeavn 290  
QY 299 ---SSGOKKQKLVPRYDVSITLWTKDLSLGGNSMTAMIAISPADINFEETLSTLRVADS 355  
DB 291 ctskkskkktdfipyrdsvltllrenlgnsrtaamvaaispadinydetlstlryadr 350  
QY 356 AKRIENHAVVNEDPNARMRELKELAQRLSKLQSSGGG-----GGAG----- 399  
DB 351 akqikcnavinedpnakivrelkeevtrklldlraqgldiiddlpllddyvgsgskylk 410  
QY 400 -----GGGGPVVEESYPDPDTLEKO-----IVSIQOPDATVKK 431  
DB 411 dfqnnkhyllasengrghfstamgsitss-pscsllssqvgltstvtisq--erimst 467  
QY 432 MSKAEIVEQLNQSEKLYRDLNQTWEKLAKEEIIHKEREAALEELGISIEK--GFVGPYH 489  
DB 468 pggaeaielkesekiliaelnetweeklrkteairmereallaeangvalredggtlgvfs 527  
QY 490 SKEMPHLVNLSDDPLLAELCLVNIKPGQTRVGNVNVQDTQAEIRLNGSKILKEHCFENV- 548  
DB 528 pkktphlvnlndplmsecllyikdgtirvgadaerrqdivlsgahikehcfirser 587  
QY 549 ----DNVTVIVNEKAAVWNGVRDKPTRLRSYRIILGDPHIFRHNHPEARAEQEO 604  
DB 588 snsggevitlpeceretyngkrvsqqlrsgnrllngkhvfrfhbpegarek- 646  
QY 605 SLLRHSVTNSQLGSPAGPHDRTLKSGSDADGSRSPSPHFRGKDSDFYARREAS 664  
DB 647 -----tpsaetpsepvdwtfagrelle 668  
QY 665 ATGLDQK-----ISHLTDDDELDFD-----VOKARAVRGLVED 701  
DB 669 k-qgldmqemeklqemellykkeeaddlileqrdldyesklqalqkvetrslaaet 727  
QY 702 NEDSDSQSSFP 712  
DB 728 teeeeeeepv 738  
  
RESULT 3  
ID AAM40034  
XX AAM40034 standard; Protein; 893 AA.  
AC AAM40034;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 3179.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX (HYSE-) HYSEQ INC.  
PA  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI59190.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4; SEQ ID NO 3179; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC Assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 893 AA;  
  
Query Match 41.2%; Score 1660.5; DB 22; Length 893;  
Best Local Similarity 46.5%; Pred No. 4.8e-118;  
Matches 350; Conservative 126; Mismatches 167; Indels 109; Gaps 15;  
  
QY 4 GGNIKVVRFPNAREIDRGAKICVRMBGNQITLTPPGAEKARKSGKTTIMDGPKAPA 63  
DB 3 gasvkavvirvpfnrsremardskclqmsgsttinvnpkpket-----pkstfs 51  
QY 64 FDRSTWSDKNAP---NYARQEDLFDQDLGVPLLDNAFKYNNCIFYAGTGGSKSYSMWG 120  
DB 52 fdysywa---htspedinyasqkvryrdigeemlqhafegynvcifaygqtagagkytmg 109  
QY 121 YGK--EHGVIPRICODMPFRINELQKKNLTCTVEVSYLEIYNERNVRLDNLNSTGNLKV 178  
DB 110 kgekqgqglipgicelfsfrindtnd-nmsysvevsymeicyervrdlnpknknrlrv 168  
QY 179 RHPSTGPYEDLAKLVRSFOEINLMDGNGKARTVAATNMNETSSRSRSHAVFTLTQK 238  
DB 169 rehpllgpyvedlslavtsyndiqldmdsgnkartvaatnmnetsrshavftltqk 228  
QY 239 WHDETQMDTEKAKISLVLDLAGSERATSGATGARLKEGAENRSLSTLGRVIAALADM 298  
DB 229 rhdaetnttekvskislvdlagseradstgagtrikeganlnkslttlgkvisalae 288  
QY 299 SSG-----KOKKNOLVPYRDSVLTLLKDSLGGNSMTAMIAISPADINFEETLSTLRVA 353  
DB 289 dsgpnknkkttdfipyrdsvltllrenlgnsrtaamvaaispadinydetlstlrya 348  
QY 354 DSAKRIKHAVVNEDPNARMRELKELAQRLSKLQSSGGGGGAGGSGGPVEESYPDP 413  
DB 349 drakqircnavinedpnknklirelkddevtrldllyagqgl-----dit 392  
QY 414 PLEKQIVSIQOPDATVKKMS-----KAEIVEQLNQSEKLYRDLNQTW 455

393 dmtalvgm-spssslsalsraasvsslherillfapggseeaierlketekiaaelnetw 451  
QY 456 EEKLAATEIHKEREAAELGELISIEK--GFVGPYHSHKEMPHLVNLSDDPLLAECILVYNI 513  
Db 452 eeklrtealrmereallaemgvmaredggtlgvfpkktphlvnlnedplmsecllyyi 511  
QY 514 KPGQTRGVNVDQAEIRLNGSKILKEHCTPEN-----VDNVVTIVPNEKAAVMVNGVR 568  
Db 512 kgdrtvrgredgerrgdvlvghfkeehcvfrsdrsgseavvtlepeceadtyngkk 571  
QY 569 IDKPTLRSLGYRIILGDFHFRNPEEAREKQESLLRHSVTNSQLGSPAPGRHDTL 628  
Db 572 vtepsilrgnrlimgkshvfrfthpeaqerqr----- 606  
QY 629 SKAGSDAGDSRSDSLPHFRGKDSDFWYARREAAASAILGLDOKISHLTDELDALFDDV 688  
Db 607 -----tpcaetpaepvdfwafqrellek-qgidmk--qemeqrlqeledgy 649  
QY 689 QKARAVRRGLVEDNEDSDSQSPFPVRDKYMSN 720  
Db 650 rrereeatylle-qqrldyesklealqkqmds 680  
RESULT 4  
ID AAY51328  
XX 17-APR-2000 (first entry)  
AC AAY51328;  
XX Human KLIMP protein.  
XX KLIMP; kinesin-like motor protein; cytosolic; anticonvulsant; human;  
KW anti-Alzheimer; anti-Parkinsonian; antidiabetic; anti-ulcerative; cancer;  
KW immunomodulatory; anti-inflammatory; anti-AIDS; antirheumatic; treatment;  
KW antiarthritic; diagnosis; neurological disorder; vesicular transport.  
XX Homo sapiens.  
XX US6013454-A.  
XX 11-JAN-2000.  
XX 28-SEP-1998; 98US-0162373.  
XX 28-SEP-1998; 98US-0162373.  
XX (INCY-) INCYTE PHARM INC.  
XX Tang YT, Corley NC, Patterson C, Guegler KJ;  
XX WPT: 2000-126064/11.  
XX N-PSDB; AAZ44744.  
XX Nucleic acid sequences encoding a human kinesin-like motor protein  
PT useful for the treatment of diseases associated with  
PT inappropriate KLIMP expression such as cancers, neurological disorders  
PT and disorders of vesicular transport -  
XX Claim 1; Fig 1A-J; 38pp; English.  
XX This invention describes a novel human kinesin-like motor protein  
CC (KLIMP) (I) which has cytostatic, anticonvulsant, anti-Alzheimer's,  
CC anti-Parkinsonian, antidiabetic, anti-ulcerative, immunomodulatory,  
CC anti-inflammatory, anti-AIDS, antirheumatic and antiarthritic activity.  
CC (I) and the protein it encodes may be used in the prevention, treatment  
CC and diagnosis of diseases associated with inappropriate KLIMP expression  
CC such as cancers, neurological disorders and disorders of vesicular  
CC transport. For example, (I) (and vectors containing (I) (iv)) and the  
CC KLIMP polypeptide may be used to treat disorders associated with  
CC decreased KLIMP expression such as cancers (e.g. lymphoma, melanoma and

CC cancers of the breast lung and prostate), neurological disorders (e.g.  
CC epilepsy, Alzheimer's disease and Parkinson's disease), disorders of  
CC vesicular transport (e.g. diabetes mellitus/insipidus, Grave's disease  
CC and gastric/duodenal ulcers), and some immune/inflammatory diseases (e.g.  
CC acquired immune deficiency syndrome AIDS), rheumatoid arthritis and toxic  
CC shock syndrome). This sequence represents the human KLIMP protein  
CC described in the method of the invention.  
XX Sequence 1103 AA;  
SQ  
Query Match 41.1%; Score 1658; DB 21; Length 1103;  
Best Local Similarity 47.3%; Pred. No. 1e-117;  
Matches 353; Conservative 120; Mismatches 159; Indels 114; Gaps 16;  
QY 4 GNIKVVVRVPFNAEIDRGAKIVRMENQITILTPPPGAEKARKSGKTIMDQPKAPA 63  
Db 3 gasvkavvrfrfnaretsqdkcvsmgnttsiinp-----kqakdapsft 51  
QY 64 FDRSWSFKNA-PNVARQEDLFQDLGVPLLDNFAFGYNNCIPAYGQTSGSGKSYMMGYG 122  
Db 52 fdysywshtstedpqfasqqvvyrdigeemlllhafegynvcifayqgtgagksytmnrg 111  
QY 123 K--EHGVIPRICODMERRINEQLQKKNLCTVEVSYLEIYNERNVRLNPNSTGNLKVRE 180  
Db 112 epqggvlpqcedlfrsvsenq--saqlsyvsvevsymelycervrlllnpkrgslrvie 170  
QY 181 HPSTGPGYVEDLAKLVVRSPQEIENLMDENGNKARTVAATNNMNTSSRSFVTLTFLQKWH 240  
Db 171 hplpgyvqdlslavtsyadiadlmdcgnkartvaatnmetssrshavftvfgqrch 230  
QY 241 DEETKMDTEKVAKISLVDLAGSERATSTGATGARLEKAGEINRSITLGLRVIAALADMSS 300  
Db 231 dgitgidsekvsislvdlagseradsgarmglkeganinksltlgkvisaladmq 290  
QY 301 GKQKKQLVPYRDSVLTLLKDSLGNSTAMIAATSPADINFEETLSTLRVADSARKIR 360  
Db 291 -kkrksdfipyrdsvtllwkenlgnsrtamaiaalspadinyeelstlryadrktqir 349  
QY 361 NHAVNEDPNARMIRELKELAQLRSKLOSSGGGGGAGG----- 400  
Db 350 cnailedpnarlirelqeearlrellmaqglasaleglkteegsvrgalpavasp 409  
QY 401 -----SGPVEESYPDDTPLEKQIVSIQOPDATVTKMSKAEIVEQLNQSEKLYRDL 451  
Db 410 pvspsptthngelapsfnt--esqi-----gpeeameriqetekiaael 454  
QY 452 NOTWEEKLAETEHKEREAAELGELISIEK--GFVGPYHSHKEMPHLVNLSDDPLLAECIL 509  
Db 455 netweeklkrtealrmereallaemgavvredggtvgvfpkktphlvnlnedplmsecl 514  
QY 510 VYNIKPGQTRGVNVDQAEIRLNGSKILKEHCTPENV-----DNVVTIVPNEKAAVMV 564  
Db 515 lyhikdgvtvrgvvdmd----ikltgqfirqehclfrsibpqpdegvvtlepecegaetyv 570  
QY 565 NGVRIDKPTLRSLGYRIILGDFHFRNPEEAREKQESLLRHSVTNSQLGSPAPGRH 624  
Db 571 ngklvtpevlksgnrlvmgmhvfrrfnheqarlerer-----gvppp--- 614  
QY 625 DRTLKAGSDAGDSRSDSLPHFRGKDSDFWYARREAAASAILGLDOKISHLTDELDAL 684  
Db 615 -----pgppsepvdwnfaqlleq-qgidikle--mekrlqdl 650  
QY 685 FDDVQKARAVRRGLVEDNE---DSDS 707  
Db 651 enqyrkekeeadllleqgrlyadsds 676  
RESULT 5  
ID AAE04316  
XX AAE04316 standard; Protein; 1103 AA.  
XX AAE04316;  
AC



XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL08011.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 18516; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1773 AA;

Query Match 40.6%; Score 1635; DB 22; Length 1773;  
 Best Local Similarity 44.5%; Pred. No. 1.2e-115;  
 Matches 367; Conservative 143; Mismatches 193; Indels 122; Gaps 20;

QY 6 NIKVVVRFPNAREIDRGAKCIVRMVGNQITLTPPGAEEKARKSGKTMGPKAFAD 65  
 DB 32 svkvavrvrpfnsrelarekskleimagattainp-----kvpntsdsvkrfnfd 83  
 QY 66 RSYWSDKNAPNARQBDLQDLQPLLDNAFKYGNICNIFAYGQTSGSKSYMMGYGKE- 124  
 DB 84 ysywshdhddadstgmyvkdigeemlqbsfgdygvcifaygqtgagksytmgrqeeq 143  
 QY 125 -HGVIPIRCODMFRINELQKKNLCTVEVSYLEIYNERVROLLNPSTKGNLKVREHPS 193  
 DB 144 qedilpmickdlitriqdtetd-dlkysvevsymelycervrdlnpknkgnlrvehpl 202  
 QY 184 TGPVEDLAKLVRSFOEINLMDENKARTVAATNNKTSRSHAVFTTLTKQWDEE 243  
 DB 203 lgpvyedslkavtdyqdhldidgnkarktaatmmetsrshavfttffqrrhdm 262  
 QY 244 TKMDTEKAKISLVDLAGSRATSTGATGARLKEGAEINSLTGLRVIAALADMSSGKQ 303  
 DB 263 tnlttekvsislvdlagseradstgaktgtrlkeganinksltlgkvisalaevasakkk 322  
 QY 304 --KKNQLVPYRDSVLTLWLLKDSLGGNSWTAMIAISPADINFEETLSTLYADSARKIKN 361  
 DB 323 ntkkadfpfyrdsealtwllrenlggnsktamaiaispadinydetlstlryadrakqivc 382  
 QY 362 HAVVNEDPNARMIRELKEELQAQLRSKLOSGGGGGAGGSGGPPVEESYPDTPLEKQIVS 421  
 DB 383 kavnedanakllrelkeelqirdllkaeg-----ievqegpdgk-----vv 425  
 QY 422 IQQPDATVKKMSKAEI-----VEQLQSEKLYRDLNQTWEKLAKE 463  
 DB 426 cekrdankdelstkvtsptksnrngsttemavdqkqasekliaelnetweekllrte 485  
 QY 464 EIHKEEALEELGISI-EKGF-VGPHSKEMPHLVNLSDDPLLAECVLVYNIRPGQTRVG 521  
 DB 486 eirvqreavfaemgavkedgltvgvfpktpkphlvalnedpnlscecllyyikegltrlg 545  
 QY 522 NVNODTQAEIRLNGSKILKECHTFENVDNVTVIPNKEAAMVNGVRIDRTPRLRSYRI 581  
 DB 546 theanvpqdqlqslgshlkehctfenkstvtllphkdailyngvrklvepevlktgrsv 605  
 QY 582 ILGFHIFRPNHPEAREAEQEQSLLRHSTVNSQLGSPAGCRHRTLSKAGSDADGDSRS 641  
 DB 606 ilgknhvfrfnpeqare-----lrdki-----eteneaeeneve---kt 641

QY 642 DSPLPHERGKSDWFFYARREASAILGLD-----QKISHLTDD-----ELDALFDVQ 689  
 DB 642 dt-----gqvdmfagcellek-qgidlkaemkkrlndleeqykreklaqdgqfeeqr 693  
 QY 690 K---AR-AVRRLGLVEDNESDSQSSFPVRDVKYMSNGTIDNFSLDYTAITWP----- 735  
 DB 694 kyearidalqkveeqstmsmysspedfheeqedvytnpnyescwtareaglaawaf 753  
 QY 736 -----GTPSRDDGDALFFGDKSKQDASNVVDYELRQQAQ 772  
 DB 754 rkwyrbqftslrddlgwnaifl-----keanaaisvelkkkvqfq 792  
 RESULT 7  
 AAU19569  
 ID AAU19569 standard; Protein; 757 AA.  
 XX  
 XX AAU19569;  
 AC  
 XX  
 DT 04-DEC-2001 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polypeptide (DITHP) #155.  
 XX  
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;  
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
 KW respiratory disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200162927-A2.  
 XX  
 PD 30-AUG-2001.  
 PF  
 PF 21-FEB-2001; 2001WO-US06059.  
 XX  
 PR 24-FEB-2000; 2000US-0184693.  
 PR 24-FEB-2000; 2000US-0184897.  
 PR 24-FEB-2000; 2000US-0184898.  
 PR 24-FEB-2000; 2000US-0184768.  
 PR 24-FEB-2000; 2000US-0184769.  
 PR 24-FEB-2000; 2000US-0184770.  
 PR 24-FEB-2000; 2000US-0184771.  
 PR 24-FEB-2000; 2000US-0184772.  
 PR 24-FEB-2000; 2000US-0184773.  
 PR 24-FEB-2000; 2000US-0184774.  
 PR 24-FEB-2000; 2000US-0184776.  
 PR 24-FEB-2000; 2000US-0184777.  
 PR 24-FEB-2000; 2000US-0184797.  
 PR 24-FEB-2000; 2000US-0184813.  
 PR 24-FEB-2000; 2000US-0184837.  
 PR 24-FEB-2000; 2000US-0184841.  
 PR 24-FEB-2000; 2000US-0185213.  
 PR 24-FEB-2000; 2000US-0185216.  
 PR 12-MAY-2000; 2000US-0203785.  
 PR 15-MAY-2000; 2000US-0204226.  
 PR 16-MAY-2000; 2000US-0204525.  
 PR 16-MAY-2000; 2000US-0204821.  
 PR 16-MAY-2000; 2000US-0204908.  
 PR 16-MAY-2000; 2000US-0205232.  
 PR 17-MAY-2000; 2000US-0204815.  
 PR 17-MAY-2000; 2000US-0204863.  
 PR 17-MAY-2000; 2000US-0205221.  
 PR 17-MAY-2000; 2000US-0205285.  
 PR 17-MAY-2000; 2000US-0205286.  
 PR 17-MAY-2000; 2000US-0205287.  
 PR 17-MAY-2000; 2000US-0205323.  
 PR 17-MAY-2000; 2000US-0205324.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Panzer SR, Spiro PA, Barville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;







Qy 357 KRKNHVVNDPNRMIRREKEELAKLRSLQSSGGGGAGGGVPEESYPDPPTLE 416  
Db 307 krivnhavvndpnariirelhevetslrmlkha-----tgsqv----- 346  
Qy 417 KQIVSIQOPDATVKKMSKAEIVQNLQSEKLYRLDNLQWEEKLAKTEIHKEREAALEEL 476  
Db 347 -----gdvqdklaesenlmkqisqtweeklvkteriqnrgqalekm 388  
Qy 477 GISIEKGFVGPYHSKEMPHLVNLSDDPLLAELCLVYNIKPGQTRVGNVNDQQAIRLNGS 536  
Db 389 gisvqas--gikvcknyylvnladpslnellvyylkdrtiligrtisgqpdiqlsgl 446  
Qy 537 KILKEHCTFENVNVVTVIPNEKAAMVWVGRIDKPTLRSGYRIILGDFHFRFNHP 594  
Db 447 giqbehcvitiedsglymepvggarcfvngaaavektpmqgdrilwgnhhrfvnsp 504

RESULT 10  
AAB63190  
ID AAB63190 standard; Protein: 503 AA.  
XX  
AC AAB63190;  
XX  
DT 26-MAR-2001 (first entry)  
XX  
XX Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.  
XX  
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;  
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
KW angiogenesis; nervous system disease; Alzheimer's disease; infection;  
KW ocular disorder; corneal infection; wound healing; skin aging;  
KW food additive; preservative.  
XX  
OS Homo sapiens.  
XX  
XX W0200061629-A1.  
XX  
XX PD 19-OCT-2000.  
XX  
XX PF 06-APR-2000; 2000WO-US09071.  
XX  
XX PR 09-APR-1999; 99US-0128694.  
XX  
XX PR 20-JAN-2000; 2000US-0176931.  
XX  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX FA (ROSE/) ROSEN C A.  
XX  
XX PI Ruben SM, Komatsoulis G;  
XX  
XX WPI; 2000-647420/62.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
XX PS Disclosure; Page 489-491; 533pp; English.  
XX

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular  
CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral  
CC ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's  
CC disease, infections caused by bacteria, viruses and fungi and ocular  
CC disorders e.g. corneal infection. The polypeptides can also be used to  
CC aid wound healing and epithelial cell proliferation, to prevent skin  
CC aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities. AAF22364 to  
CC AAF22372 and AAB63133 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 503 AA;

Query Match 31.7%; Score 1276; DB 21; Length 503;  
Best Local Similarity 50.8%; Pred. No. 5.3e-89;  
Matches 277; Conservative 74; Mismatches 136; Indels 58; Gaps 11;

Qy 58 GP-KAFADFYSWFDKNA-PNYARQEDLFQDLGVPLLDNAFKGYNNCIFAYGQTGSGKS 115  
Db 9 gpaevfaydhcfwmdesvkekyagqdlvfkclgenilqnaifgynacifaygqtgsgks 68  
Qy 116 YSMGKGKEHGVIPIRICQDMFRR-----NELQKDKNLCTVEVSYLEIYNERNVRLDLP- 170  
Db 69 ytmngtdqpgliprlcsglferitqkegneesgfk-----vevsymelynekvrdllqk 123  
Qy 171 STGNLKVREHPSGYPYEDLAKLVRSFOEINLMDGNCARTVAATNMNMTSSRSNAV 230  
Db 124 gsrqtklvrehsvigpyvvgdlskavtsykieslmsegnskrtvvaatnmneessrshav 183  
Qy 231 FTLTQKWHDEETKMDTEKVAKISLVDLGSEATSTGATGARLKEAGAEINRSLSTLGR 290  
Db 184 fkitlthlyvksqtsgekvgsxlvdxgeratkgaaagdrilkegnslnksltlgl 243  
Qy 291 VIALADMSGQKKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAAISPADINFEETLSTL 350  
Db 244 visaladqsagk-sxnkfpyrdsvlwllkdsiggnskxamvatvspaadnydetlsl 302  
Qy 351 RYADSAKRIKNHVVNDPNRMIRREKEELAKLRSLQSSGGGGAGGGVPEESYP 410  
Db 303 ryadrakhivnhavvndpnariirdirevekeklregitka----- 343  
Qy 411 PDTPLEKQIVSIQOPDATVKKMSKAEIVQNLQSEKLYRLDNLQWEEKLAKTEIHKERE 470  
Db 344 -----eamkspeldrieeseekligemtvtweeklrkteelaqerg 384  
Qy 471 AALEBELGISIEKGFVGPYHSKEMPHLVNLSDDPLLAELCLVYNIKPGQTRVGNVNDQQA 530  
Db 385 kqleslgisldqs--gikvgddkcfvlnadpalnellvyyke-htligsansq--- 438  
Qy 531 IRLNGSKILKEHCTFE-NVDNVVTVIPNEKAAMVWVGRIDKPTLRSGYRIILGDFHIF 589  
Db 439 igicmgilpbehcliditseggmiltpqkntrrtfngssvspiglhghdrilwgnhff 498  
Qy 590 RFNHP 594  
Db 499 rlnhp 503

RESULT 11  
AAB63190  
ID AAB63190 standard; Protein: 421 AA.  
XX  
AC AAB63190;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 6751.  
XX  
KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukaemia.	
XX	Homo sapiens.	
OS		
XX	WO200153312-A1.	
PN		
XX	26-JUL-2001.	
PD		
XX	26-DEC-2000; 2000WO-US34263.	
PF		
XX	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0532317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
PR	29-NOV-2000; 2000US-0727344.	
XX	(HYSE-) HYSEQ INC.	
PA		
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang JT, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou F, Goodrich R, Drmanac RT;	
XX		
DR	WPI; 2001-442253/47.	
DR	N-PSDB; AAI60976.	
XX		
XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
PT		
XX	Example 2; SEQ ID NO 6751; 10078pp; English.	
PS		
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
CC	the encoded polypeptides (AAI38642-AAI42213) with nootropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression.	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SQ	Sequence 421 AA;	
	Query Match 31.0%; Score 1251; DB 22; Length 421;	
	Best Local Similarity 61.1%; Pred. No. 3.3e-87;	
	Matches 245; Conservative 68; Mismatches 64; Indels 24; Gaps 6;	
QY	4 GGNIKVVVRFPNAREIDRGAKCIVRMENQOTILTTPPGAEKARKSGKTMIDGPKAFA 63	
DB	9 gasvkvavrvrpfnsremrsdskliqmsgstttivnpkpket-----pkfsfs 57	
QY	64 FDRSWSFDKNAP---NYARQEDLFODLVPLLDNAFGYNNCFAYGOTSGSGKSYSMMG 120	
DB	58 fdysyws--htspedinyasqkvqrldigeemlqatageynvcifay9qrgagksytimg 115	
QY	121 YGK--EHGVIPRICODMFRFRELQKKNLACTVEVSYLEYNERVDRLLNPSTKGNLKV 178	
DB	116 kqeqdggglipqlcdlfsrindttnd-nmsysvevsymeicyercervrldllnpknknlrv 174	

QY	179	REHPSGTGYPVEDLAKLVVRVSFOETIENLMDEGNKARTVAATNMNETSSSHAVFTLTQK	238
DB	175	rehpllgpyvedsklavtysyndiqldmsdgnkartvaatnmnetsrshavfniftqk	234
QY	239	WHEDETKMDTEKVAKISLVDLAGSERATSTGATGARLKEGAEINRSLSTLGRVIAALADM	298
DB	235	rhdaetnlttekvsksislvdiagseradstgagtrlkeganinkslctllgkvisalaem	294
QY	299	SSG-----KQKNQLVYPYRDSVLTWLLKDSLGGSMTAMIAAISPADINFEETLSTLYA	353
DB	295	dgpnknkkkktfdpyrdsvtllrenlgnnartamvaalspadinydetlstrya	354
QY	354	DSAKRIKNHAVVNEPDNARMIRELKEELAQLRSKLQSSGGG	394
DB	355	draklrcnavinedpnknklrelkdevtrirdliyaqglg	395
	RESULT 12		
	ABB61704		
ID	ABB61704 standard; Protein; 1174 AA.		
XX	AC	ABB61704;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 11904.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
XX	OS	Drosophila melanogaster.	
XX	PN	WO200171042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US09231.	
XX	PR	23-MAR-2000; 2000US-191637P.	
XX	PR	11-JUL-2000; 2000US-0614150.	
XX	PA	(PEKE ) PE CORP NY.	
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	XX	WPI; 2001-656860/75.	
XX	DR	N-PSDB; ABL05807.	
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
XX	PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
XX	PT	interactions -	
XX	PS	Disclosure; SEQ ID NO 11904; 21pp + Sequence Listing; English.	
XX	CC	The invention relates to an isolated nucleic acid detection reagent	
XX	CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
XX	CC	useful in developmental biology and in elucidating cell signalling and	
XX	CC	cell-cell interactions in higher eukaryotes for the development of	
XX	CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
XX	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
XX	CC	sequences (ABL01840-ABL16175) and the encoded proteins	
XX	CC	(ABB57737-ABB72072).	
XX	CC	The sequence data for this patent did not form part of the printed	
XX	CC	specification, but was obtained in electronic format directly from WIPO	
XX	CC	at ftp.wipo.int/pub/published_electronic_sequences.	
XX	XX	Sequence 1174 AA;	
	Query Match 28.0%; Score 1128; DB 22; Length 1174;		
	Best Local Similarity 36.5%; Pred. No. 4.1e-77;		
	Matches 286; Conservative 125; Mismatches 267; Indels 106; Gaps 21;		



Db 582 lrqlaerereisraqk-----smeklkkaedqrkselrvlkriglael-----ta 629  
QY 490 SKEMPHLVNLDPLAEGVLVNIKPGQTRVG-----NVNQDTQAEIRLNGSKILKEHCT 544  
Db 630 eqkacqlvntadpilsqgtifylpqlgrlrglrpssgssqpdvldgplvalghcs 689  
QY 545 FENV--DNVVTIVPNEKAAVMVNGVRIDKPTRLRSGYRIILGDHFIFRNHP---EARA 599  
Db 690 ieherggklyvpssedfetyvngelkrrqlfghdrivigshyfrisnfcsgrgka 749  
QY 600 EROEQSLLRH-----SVTNSQLGSPAPGRHRTLSKAGSDADGSDSDS-----PLPHFR 649  
Db 750 dhpvdqlahqellkqegqlrseleakraaltkieqeraghardfeerlqcleleqfk 809  
QY 650 GK--DSDFYARRAASAILLDQKISHLTDELDALFDVQKA---RAVRRLGIVEDNDS 705  
Db 810 ykcnsemleterda---lalaqqgthtprlhedaavstpaqkstlledlqrimlpsees 865  
QY 706 DSQSFVPVRD-----KYSNGSIDNFSL 728  
Db 866 lhktqlmvkeatqrcrqlplplefrtqtqpdqgl 900

## RESULT 14

AAW03691  
ID AAR57365 standard; Protein; 955 AA.

AC AAR57365;

DT 08-MAR-1995 (first entry)

XX K39 polypeptide of Leishmania chagasi.

XX Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;  
XX Leishmania chagasi.

OS Leishmania chagasi.

XX WO9416331-A.

XX 21-JUL-1994.

XX 10-JAN-1994; 94WO-US00324.

XX 15-JAN-1993; 93US-0006676.

XX (IASY-) IASYS CORP.

XX Read SG;

XX WPI; 1994-249402/30.

XX N-PSDB; AAQ70152.

XX Diagnosis of Leishmaniasis - by determining the presence of  
XX antibodies that bind to a K39 repeat unit antigen

XX Disclosure; Page 12-15; 28pp; English.

XX The K39 polypeptide comprises a number of repeated units (described  
XX in AAR57366). Detection of antibodies directed against this repeated  
XX unit in a patients sample is indicative of Leishmaniasis. The  
XX antigenic repeat unit can itself be used as a vaccine to protect  
XX against infection by a leishmania parasite.

XX Sequence 955 AA;

Query Match 20.28; Score 814; DB 15; Length 955;

Best Local Similarity 28.18; Pred. No. 3.2e-53;

Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY 7 IKVVVRVRFNARE---IDRGAKCIVRMEGNQITLTPP-----PGAFKARKSGKTMD 57

Db 13 VKSVVRVPLHERENNAPGCTKVTVAQAAAVVTKVVGSSNNSSGAESMGTRARRVAQD 72  
QY 58 GPKAFAPDRSYWSF---DKNAPNARYARQEDFLQDLGVPLLDNAPKGYNNCIFYAQOTSGSK 114  
Db 73 ---fqdhvfwsvetpdacatpadvrtgypivqahfgdfnscfayqgtsgk 128  
QY 115 SYSMMG-----YKHEGVIPRICODMPRRINELOKKNLCTCTVEVSYLEIYNERVRL 168  
Db 129 tytmngadvsalsgegngvtpriciefarkasveagqhsrwivelgyveyvnervsdl 188  
QY 169 NPSTGK-----NLKYREHPSTGPIYVEDLAKLVVRSEFQEIENLMDGKNKARTVAATNM 220  
Db 189 gkrkgvkgggeeveydvrehpsrgvfliegqlrvevgsladvvrlieigngvrhtastkm 248  
QY 221 NPTSSRSHAVFTLTL---TQKHDEETKMDTEKVAISLVLDLAGSERATSTGATGARLK 276  
Db 249 ndrsrshaimlllreertmttksetirtagksrnmnlvdlagservaqsgvqgfk 308  
QY 277 EGAEINRSLSTLGRVIAALADMS--GKOKKQOLVYPRDSVLTWLLKDSLGNSMTAMTAA 335  
Db 309 eathlnslttlgrvldmadmatkgakqsvapfrdskitfildksglgnsktftmat 368  
QY 336 ISPADINPEETLSTLYADSAKRIKHNHVVNEDPNARIRELKEELAQLRSKLSQSSGGG 395  
Db 369 vspalnlyeetlstlryasaradivnvagvnedprarrireleeqmedmrqam----- 421  
QY 396 GGAGSGGGPVVE-----SYPTPTPLEK 417  
Db 422 --aggdpayvselkklalleseagkraadlqalererehngvqerllrateakeles 479  
QY 418 QIVSTQQ-----PDATVKMKSKAEIVEQL 441  
Db 480 raalqeamtatrrqdkmqalnrlkeeqarkerellkemakkaalskvrirrkdaeia 539  
QY 442 NOSEKLYRDLNQTWEKLAKEEIIHKEREAALFELGISIEKGFVGPYHSEKMPHLVNLSD 501  
Db 540 serek1-----estvaqlereqrevaldal-----qthqrkiquealesse 581  
QY 502 -----DPLLAELVYNIKPGQTRVGNVNDTQAEIRLNGSKILKEHCTFENVVNVTV 555  
Db 582 rtaaedqllq--ltelqsertqlsqvvdretrlt-rdlqriqyeygetelardvalca 638  
QY 556 PNE-----KAAVMVNGVRIDKPTRLRSGYRIILGDHFIFRNHPEARAERQESLLRHS 610  
Db 639 agemearyhaavfhlqtllelatewedalr-----eralaeadaaae-- 682  
QY 611 VTNSQLGSPAPGRHRTLSKAGSDADGSDSDSPLPHFRGKSDWFYARREASAILGLD 670  
Db 683 -----ldaaastsqu-----aresaceritsle 705  
QY 671 QKISHLTDELDALFDVQKARAVRGLYVEDNEDSDSQSFVPVRDKYMSNGITDNFSLDT 730  
Db 706 qqlre-seeraaelasqleataaaksaqcdrentatleqlres-----earaael 757  
QY 731 AITMPTGTPSRDSDGDLFFGDKKSKQDASNV-----DVEELRQQAQMEALKTKAR 781  
Db 758 asqleataaa-----kmsaqcdrentatleqlrdseeraaelasqlestaak 807

## RESULT 15

AAW03691

ID AAW03691 standard; Protein; 955 AA.

XX AAW03691;

XX 09-MAR-1997 (first entry)

XX Leishmania chagasi K39 antigen.

XX Leishmania chagasi; acidic ribosomal antigen; Lcp0;

XX epitope; K39.

XX



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 04:39:26 ; Search time 25.38 Seconds  
(without alignments)  
754.518 Million cell updates/sec

Title: us-09-235-416-1

Perfect score: 4030

Sequence: 1 MSGGNIKVVVRPFNARE.....ELRQQAQMEALKAKQEF 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1658	41.1	1103	3	US-09-162-373-1
2	1658	41.1	1103	4	US-09-467-946-1
3	1311.5	32.5	1375	4	US-09-722-139-2
4	1005.5	25.0	563	4	US-09-718-841-2
5	991	24.6	359	4	US-09-722-139-4
6	814	20.2	955	1	US-08-006-676B-1
7	814	20.2	955	1	US-08-282-845-2
8	814	20.2	955	5	PCT-US94-00324-1
9	812.5	20.2	955	2	US-08-428-414A-3
10	747	18.5	357	4	US-09-718-841-4
11	673	16.7	1057	4	US-09-541-782-10
12	664.5	16.5	1388	4	US-09-572-191-2
13	640.5	15.9	409	4	US-09-572-191-6
14	633	15.7	1184	4	US-09-541-782-2
15	625	15.5	411	2	US-08-713-815A-3
16	625	15.5	441	2	US-08-713-815A-4
17	610.5	15.1	1066	4	US-09-541-782-8
18	591	14.7	375	4	US-09-572-191-4
19	587.5	14.6	1038	4	US-09-541-782-4
20	579	14.4	1073	4	US-09-541-782-6
21	541	13.4	299	4	US-09-621-233-2
22	513.5	12.7	790	4	US-09-723-153-2
23	464	11.5	326	4	US-09-721-383-4
24	464	11.5	787	4	US-09-721-383-2
25	457.5	11.4	338	4	US-09-723-153-4
26	304.5	7.6	1201	4	US-09-098-901-2
27	258	6.4	205	4	US-09-637-481-2

28	195	4.8	1612	3	US-08-545-860D-48	Sequence 48, Appl
29	195	4.8	1612	5	PCT-US94-04496-48	Sequence 48, Appl
30	185.5	4.6	1829	4	US-09-157-420-1	Sequence 1, Appl
31	143	3.5	3248	1	US-08-353-700-1	Sequence 1, Appl
32	143	3.5	3248	5	PCT-US95-16216-1	Sequence 1, Appl
33	142	3.5	2482	1	US-08-328-254-6	Sequence 6, Appl
34	140	3.5	1093	5	PCT-US93-03077-1	Sequence 1, Appl
35	138	3.4	1354	3	US-08-685-871-2	Sequence 2, Appl
36	137	3.4	976	4	US-09-104-324B-4	Sequence 4, Appl
37	136	3.4	1388	2	US-08-685-576-4	Sequence 4, Appl
38	135.5	3.4	700	1	US-07-720-589-2	Sequence 2, Appl
39	135.5	3.4	700	5	US-08-785-190-2	Sequence 2, Appl
40	135.5	3.4	700	5	PCT-US92-05539-2	Sequence 2, Appl
41	134.5	3.3	859	4	US-09-199-637A-281	Sequence 281, App
42	133.5	3.3	700	4	US-08-235-836C-66	Sequence 66, Appl
43	133	3.3	1388	2	US-08-685-576-1	Sequence 1, Appl
44	132.5	3.3	1702	4	US-08-296-791-5	Sequence 5, Appl
45	132.5	3.3	1702	5	PCT-US95-10661A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-09-162-373-1  
; Sequence 1, Application US/09162373  
; Patent No. 6013454  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN  
; FILE REFERENCE: PF-0593 US  
; CURRENT APPLICATION NUMBER: US/09/162,373  
; CURRENT FILING DATE: 1998-09-28  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 1103  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: 1281811  
US-09-162-373-1

Query Match	41.1%	Score 1658;	DB 3: Length 1103;
Best Local Similarity	47.3%	Pred. No. 5.1e-131;	
Matches	353;	Conservative 120;	Mismatches 159;
		Indels 114;	Gaps 16;
Qy	4	GNIKVVVRPFNAREIDRGAKCIVRMENGTIILTPPPGAEEKARKSKGKTMDGPKAFA	63
Db	3	GASVQVAVRPFNARETSQDAKCVVSMQNTTSINP-----KQSKDAKPKSFT	51
Qy	64	FDRSTWSFDKNA-PNYAKQEDLFQDLGVPLDINAQKYNVICFAYGQTSGSKSYMMGVG	122
Db	52	FDYSYWSHTSTEDPOFASQQQVYRDIQBEMLLHAFEGYVNCIFAYGQTCGAGSKSYMMGRQ	111
Qy	123	K--EHGVTPRICODMFRINELQDKNLCTCTVEVSYLEIYNERNVRLDNLNPSKGNLKYRE	180
Db	112	EPGQGGIVPQCEDEDFSRVSENG-SAQLSYSVEVSMEIYICERVRLDNLNPKSRGSLRVRE	170
Qy	181	HPSTGPYVEDLAKLVRSFOEINLDEGNKARTVAATNMNETSSRSNAVFTLTITQKWH	240
Db	171	HPILGPYQDLSKLAITYADIADLMDCGNKARTVAATNMNETSSRSNAVFTIVTQRCH	230
Qy	241	DEETKMDTEKVAKISLVLDLAGSERATSTGATGARLKEGAEINRSLSLTGRVTAALADMSS	300
Db	231	DQLTGLDSEKYSKISLVLDLAGSERADSSGARGMGLKEGANINKSLTTLGKVISALADMOS	290
Qy	301	GQKNQLVPPYRDSVLTWLLKDSLCGNSMTAMIAISPADINFETLSLTRYADSAKRITK	360

Db 291 -KKRSDFIPYRDSVLTWLLKENLGNSTAMIAALSPADINYEETLSTLAYADRTKQIR 349  
QY 361 NHAVNEDPNARMIRELKEELAQSLKSSGGGGGAGG----- 400  
Db 350 CNAIINEDPNARLIRELQEEVARLRELLMAOGLSASALEGLKTEGSRGALPAYSPPA 409  
QY 401 -----SGGPVEESYPDPDPLEKQIVSIQQPDATVKMKSKAEIVBQLNQSEKLYRDL 451  
Db 410 PVSPPSTTHNGELEPSPNPT--ESQI-----GPEAMERLQETEKIIAEL 454  
QY 452 NOTWEEKLAKEEIHKEERAALEELGISTEK--GFVGPYHSEMPHLYNLNDDPILLAEC 509  
Db 455 NETWEEKLRKTEALRMEREAALLAEMGVAVREDGGTVGVFSPKKTPLHVLNLEDPLMSECL 514  
QY 510 VYNIKPGQTRVGNVODTQAEIRLNGSKILKEHCTFENV-----DNVVTIVNEKAAVMV 564  
Db 515 LYHIKGVTRVGOVDM-----IKLTQFTREOHCFLRSIPQDPGEVVVILECEGAETV 570  
QY 565 NGVRIDKPTRLRSYRIILGDFHIFRNPHEPARAQBSLLRHSVTNSQLGSPAPGRH 624  
Db 571 NGKLVTPLVLKSGNRIVMGKRVFRNHPHQARLER-----GVPPP----- 614  
QY 625 DRTLKAGSDADGSDRSPLPHFRGKSDWYARREASAILGLDOKISHLTDDDELDA 684  
Db 615 -----PGPPSEPDVWNPFAQKELLEQ--QGIDIKLE--MEKRLQDL 650  
QY 685 FDDVQKARAVRGLVEDNE---DSDS 707  
Db 651 ENQYRKEEKEADLLLEQRLYADSDS 676  
RESULT 2  
US-09-467-946-1  
; Sequence 1, Application US/09467946  
; Patent No. 6248594  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN  
; FILE REFERENCE: PF-0593 US  
; CURRENT APPLICATION NUMBER: US/09/467,946  
; CURRENT FILING DATE: 1999-12-21  
; EARLIER APPLICATION NUMBER: 09/162,373  
; EARLIER FILING DATE: 1998-09-28  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 1103  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: -  
; OTHER INFORMATION: 1281811  
US-09-467-946-1  
Query Match 41.1%; Score 1658; DB 4; Length 1103;  
Best Local Similarity 47.3%; Pred. No. 5.1e-131;  
Matches 353; Conservative 120; Mismatches 159; Indels 114; Gaps 16;

QY 4 GGNKVVVRPFPNAREIDRGAKEIVRMENQITLTTPPGAEEKARKSGKTMDGPKAF 63  
Db 3 GASVAVVRPFPNARETSQDAKCVWSQGNNTSIINP-----KOSKDAKSF 51  
QY 64 FDRSVSWFDKNA-PNARQEDLFQDLGVPLLDNAFKYNNCIFAYGTGSGKSYSMGYG 122  
Db 52 FDSYTSWHSSTEDPQFASQOQYVRDIEGEMLLHAFEGYNVCIFAYGTGAGKSYTMGRQ 111  
QY 123 K--EHGVIPRICQDMFRINELQDKNLTCTVSVSYLBIYNERVRLNLPSTKGNLYRE 180  
Db 112 EPGQGIIVPOLCEDLFSRVSNO--SAQLSVSVSYMEIYCERVRLDNLNPKSRGSLRYRE 170

QY 191 HPSTGPPYVEDLAKLVRSFQRIENLMDGNKARTVAATNNMNETSSRSRAVFTLTITOKWH 240  
Db 171 HPILGPYQDLSKLAIVSYADIADLMDCGNKARTVAATNNMNETSSRSRAVFTIYFTQCH 230  
QY 241 DEETKMDTEKVAKISLVDLASERATSTGATCARLKEGAENRSLSITLGRVIAALADMS 300  
Db 231 DQTLGLDSEKYSKISLVDLASERADSSGARGMLKEGANINKSLITLGVISALADMS 290  
QY 301 GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAAISPADINFEETLSTLAYADSAKRIK 360  
Db 291 -KKRSDFIPYRDSVLTWLLKENLGNSTAMIAALSPADINYEETLSTLAYADRTKQIR 349  
QY 361 NHAVNEDPNARMIRELKEELAQSLKSSGGGGGAGG----- 400  
Db 350 CNAIINEDPNARLIRELQEEVARLRELLMAOGLSASALEGLKTEGSRGALPAYSPPA 409  
QY 401 -----SGGPVEESYPDPDPLEKQIVSIQQPDATVKMKSKAEIVBQLNQSEKLYRDL 451  
Db 410 PVSPPSTTHNGELEPSPNPT--ESQI-----GPEAMERLQETEKIIAEL 454  
QY 452 NOTWEEKLAKEEIHKEERAALEELGISTEK--GFVGPYHSEMPHLYNLNDDPILLAEC 509  
Db 455 NETWEEKLRKTEALRMEREAALLAEMGVAVREDGGTVGVFSPKKTPLHVLNLEDPLMSECL 514  
QY 510 VYNIKPGQTRVGNVODTQAEIRLNGSKILKEHCTFENV-----DNVVTIVNEKAAVMV 564  
Db 515 LYHIKGVTRVGOVDM-----IKLTQFTREOHCFLRSIPQDPGEVVVILECEGAETV 570  
QY 565 NGVRIDKPTRLRSYRIILGDFHIFRNPHEPARAQBSLLRHSVTNSQLGSPAPGRH 624  
Db 571 NGKLVTPLVLKSGNRIVMGKRVFRNHPHQARLER-----GVPPP----- 614  
QY 625 DRTLKAGSDADGSDRSPLPHFRGKSDWYARREASAILGLDOKISHLTDDDELDA 684  
Db 615 -----PGPPSEPDVWNPFAQKELLEQ--QGIDIKLE--MEKRLQDL 650  
QY 685 FDDVQKARAVRGLVEDNE---DSDS 707  
Db 651 ENQYRKEEKEADLLLEQRLYADSDS 676  
RESULT 3  
US-09-722-139-2  
; Sequence 2, Application US/09722139  
; Patent No. 6355471  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1055  
; CURRENT APPLICATION NUMBER: US/09/722,139  
; CURRENT FILING DATE: 2000-11-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Human  
US-09-722-139-2  
Query Match 32.5%; Score 1311.5; DB 4; Length 1375;  
Best Local Similarity 39.1%; Pred. No. 1.3e-101;  
Matches 309; Conservative 117; Mismatches 241; Indels 123; Gaps 13;

QY 6 NIKVVRPFPNAREIDRGAKEIVRMENQITLT-----PPGAEEKARKSGKTMDGPKA 61  
Db 3 SVKAVVRPFPNMRKLEAKFIQMEKSKTITNLKIPPEGTGDGSRERTKT----- 55  
QY 62 FADRSWSFDKNA-PNARQEDLFQDLGVPLLDNAFKYNNCIFAYGTGSGKSYSMGY 121  
Db 56 FTYDFSYSADTKSPDVSQEMVFKTLGTDDVKSFAEGYNACVAFYGTGSGKSYTMGN 115



Qy 122 GKEGVIPRICODMPRRINELOKKNLCTCTEVSVLEYIYNERVROL--NPSTKGNLKV 179  
Db 116 SGDSLIPRICEGLFSRINETRWDEASFRTEVSLEYIYNERVROLARKSKSTENLVR 175  
Qy 160 EHPSTGPPVEDLAKLVVRSFOEINLMDGKARTVAATNMNETSSRSRHAVFTLTQKW 239  
Db 176 EHPKGGPVYDLSKHLVQYIGDVEELMDAGNINRTATGMNDVSSRSHAIPTIKTQAK 235  
Qy 240 HDEETKMDTEKVAKISLVLDAGSERATSTGATGARLKEGAENRSLSTIGRVAALADMS 299  
Db 236 PDSE--MPCETVSKIHLDVLAGSERADATGATGVRLKEGNINKSLVTL----- 282  
Qy 300 SKQKKNQOLVYRDSVLTWLLKDSLGGSMTAMIAAISPADINFEETLSTLYADSAKRI 359  
Db 283 -GAKKQVFPYRDSVLTWLLKDSLGGSMTAMIAAISPADINFEETLSTLYADSAKRI 341  
Qy 360 KHAVVNEDPNARMTRELKEELAOLRSKLOSGGGGGAGGSGGPVSEYPPDTLEKQI 419  
Db 342 INKPTINEDANVKLIRELRAEIAIUKTLAOG-----NOI 376  
Qy 420 VSIQPDATVTKMSKAEIVQEQNLQSEKLYRDLNQTWEKLAKEEIHKEERAAELGIGS 479  
Db 377 ALLDSPTA-----LSMEELQONEARVOELTKETNKNWNETQNLKKEOTLAKRKEGI- 428  
Qy 480 IEKGVGYPYHSEMPHLYNLSDDPLLAELCLVYNIKPGQTRGVNVNODTQAEIRLNGSKIL 539  
Db 429 -----GVYLDSELPHLIGIDDLSTGIILYHLKEGQTYVGRDDASTEQDVLHGLDLE 482  
Qy 540 KEHCTFENVDNVVTIVPNEKAAMVNGVRIDKPTRLRSYRILGDFHFRFNHPPEEA-- 597  
Db 483 SEHCIFENIGTVTLIPLSGSCSYNGVQIVTEATHLNGAVJLLGRTNMFRNHPKRAAK 542  
Qy 598 -RAERQESLLRHSYTNLSQSPAGRPGRDRTLSKAGSDADGSDRSPLPHPRGKDSOWF 656  
Db 543 LREKRSGLLSFSLSMTDLKSRNLSAVMLYNGLEPERQORRELEKLESKRILIEEM 602  
Qy 657 YARRAASAILGLOOKISHLTDELDFDDVO---KARAVRGLVEDNEDSDSSPPV 713  
Db 603 EEKQSDKA-----ELRMQOEVEQTEKETEIVQLQIRKQEEELKRSEFHI 648  
Qy 714 RDKYMSNGTIDNFSLDITAMPTGRSDDDGDALFFGDKSKQDASNDVDELROQQAQM 773  
Db 649 ENKL-----KDLAEEKEFEERLEQQ---- 671  
Qy 774 EBALKTAQKE 783  
Db 672 ETELQKKRQE 681

RESULT 4  
US-09-718-841-2  
; Sequence 2, Application US/09718841  
; Patent No. 633184  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 633184el motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1051  
; CURRENT APPLICATION NUMBER: US/09/718,841  
; CURRENT FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (16)...(37)  
; OTHER INFORMATION: Xaa - any amino acid  
; NAME/KEY: VARIANT

; LOCATION: (391)...(412)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-718-841-2  
  
Query Match 25.0%; Score 1005.5; DB 4; Length 563;  
Best Local Similarity 39.0%; Pred.No. 2e-76;  
Matches 238; Conservative 87; Mismatches 224; Indels 61; Gaps 11;  
  
Qy 6 NIKVVVVRVPPNAREIDRGAKCIVRMEGNOTILTPPPGAEEKARKSGKTMIDGPKAFAD 65  
Db 3 NVQAVVRVPLSKXXXXXXXXXXXXXXXXXXVNDNRPDGFGDSREKVM---AFGD 58  
Qy 66 RSYWFDKNAPYAROEFLDGLVPLLDNAPKFNCCIFAYGQTGSGKSYGMMGYGRKH 125  
Db 59 YCWSVNPEDPOYASQDVVFDLGMVLSGVAKYGNICLFAYGQTGSGKTYTMLGTPASV 118  
Qy 126 GVIPRICODMPRRINELOK---KNLTCTEVSVLEYIYNERVROLNPSYTKG---NLKV 179  
Db 119 GLTPRICEGLFVR---EKDCASLPSSCRKIKVFLEIYNERVROLKQSGQKKSITLVR 174  
Qy 180 EHPSTGPPVEDLAKLVVRSFOEINLMDGKARTVAATNMNETSSRSRHAVFTLTQKW 239  
Db 175 EHPENGPYVGLSOHVYNYKQVIOLEEGIANRRTAATHVHEASSRSHAIPTIHYTQA- 233  
Qy 240 HDEETKMDTEKVAKISLVLDAGSERATSTGATGARLKEGAENRSLSTIGRVAALADMS 299  
Db 234 -TLENNLPSEMAKINLVLAGSERADPSYCKD-RIAGANINKSLVTLGIVISTLAQNS 291  
Qy 300 ---SCQKKN--QLVYRDSVLTWLLKDSLGGSMTAMIAAISPADINFEETLSTLYAD 354  
Db 292 QVFSQCSLNSSYIPYRDSVLTWLLKDSLGGSMTAMIAAISPADINFEETLSTLYAD 351  
Qy 355 SAKRIKNAHVNEDPNARMTRELKEELAOLRSKLOSGGGGGAGGSGGPVSEYPPDTP 414  
Db 352 SAKNIINPRVNDANLKLIRELEERLKALLS-----FE 389  
Qy 415 LEKQIVSTQPDATVTKMSKAEIVQEQNLQSEKLYRDLNQTWEKLAKEEIHKEERAALE 474  
Db 390 LXXXXXXXKXXXXXXXXXXIDQLT-----KDWTKW-----NDWQALME 432  
Qy 475 ELGISTEGFVGPYHSEMPHLYNLSDDPLLAELCLVYNIKPGQTRGVNVNODTQAEIRLN 534  
Db 433 HYSVDINRRRAGVVIDSSLPHLMALEDVLTGVVLYHLKEGTYKIGRIDSDQEQDIVLQ 492  
Qy 535 GSKILKEHCTFENVDNVVTIVPNEKAAMVNGVRIDKPTRLRSYRILGDFHFRFNHP 594  
Db 493 GOWIERDCHTITSAGCVVVLPRGARCIVNGREVTASCRLTQGAIVITLGAQKFRFNHP 552  
Qy 595 EEARAERQHQ 604  
Db 553 AEAAVLRQR 562

RESULT 5  
US-09-722-139-4  
; Sequence 4, Application US/09722139  
; Patent No. 6355471  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1055  
; CURRENT APPLICATION NUMBER: US/09/722,139  
; CURRENT FILING DATE: 2000-11-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Human  
; NAME/KEY: VARIANT  
US-09-722-139-4

Query Match	24.6%;	Score 991;	DB 4;	Length 359;
Best Local Similarity	57.0%;	Pred. No. 1.7e-75;		
Matches 208;	Conservative 48;	Mismatches 89;	Indels 20;	Gaps 5;

Y	6	NIKVVVRPFNAREIDRGAKCIVRMENQNTILT---	PPGAEEKARKSGKTI	MDGPKA	61		
b	3	SVKVVAVRPMNRREKDLAKFIOMEKSKTITINLK	IEGGTGD	SGRERTKT	55		
Y	62	FADRSWSDKKNAPYAREDFODLGVPLLDNAFKG	YNNCIFAYGQTGSGKSY	SMMGY	121		
b	56	FTYDFSYSADTSPDVSQEMVKILGTDVVKSA	EGYNACVAFAYGQTGSGKSY	TMMGN	115		
Y	122	GKEHGVIPRICODMFRINLQDKNLTCTVEVSY	LEINERVDLL--NPSTKGNL	KVYR	179		
b	116	SGDSLIPRICEGLFSRINETRWDEASFEVSY	LEINERVDLLRKRKSK	TFLRYR	175		
Y	180	EHPSTGYPVEDLAKLVYRSFQETENLMDGK	KARTVAATNMN	TSKSHAVFTLT	TQKW	239	
b	176	EHPKEGYPVEDLSKHLVQNYGDVEELMDAG	NIINRTNATGNDV	SSRSHAI	FTIKFTQAK	235	
Y	240	HDEETKMDTKVAKISLVLDLAGSERATP	STGATGARKKEGAE	INRSLS	TLGRVTAAL	ADMS	299
b	236	FDSE--MPCETVSKHLVDLAGSERADAT	GATGVLKKEGNI	KNKSLV	TLGNVISAL	ADLS	293
Y	300	SG-----KQKNQLVYRDSVLTLLKDS	LGGSMTAMIAA	ISPADIN	FEETLST	LYAD	354
b	294	QDAANTLAKKQVFPYRDSVLTLLKDS	LGGSMTAMIAA	ISPADIN	FEETLST	LYAN	353
Y	355	SAKRI	359				
b	354	RAKNI	358				

RESULT 6

US-08-006-676B-1

Sequence 1, Application US/08006676B

Patent No. 5411865

GENERAL INFORMATION:

APPLICANT: Reed, Steven

TITLE OF INVENTION: Diagnosis of Leishmaniasis

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jeffrey B. Oster

STREET: 8339 SE 57th Street

CITY: Mercer Island

STATE: Washington

COUNTRY: USA

ZIP: 98040-4906

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORD for Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/006.676B

FILING DATE: 15-JAN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oster, Jeffrey B.

REGISTRATION NUMBER: 32,585

REFERENCE/DOCKET NUMBER: REED-4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 232 7845

TELEFAX: (206) 236 0205

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 955 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-006-676B-1

Query Match	20.2%;	Score 814;	DB 1;	Length 955;
Best Local Similarity	28.1%;	Pred. No. 6.9e-60;		
Matches 253;	Conservative 125;	Mismatches 292;	Indels 230;	Gaps 24;

QY	7	IKVVVRVPPFNARE--IDRGAKCIVRMENQNTILT	PP-----PGAEKARKSGKTI	MD	57																									
DB	13	VKVSVRVPLNERENNAPEGTKVTVAAKAAAVTV	KVLGGSNNSGAESMG	TARRVAQD	72																									
QY	58	GKPAFAPDRSTWSF---DKNAPNTARQED	LFODLGVPLLDNAFKG	YNNCIFAYG	QTSGGK	114																								
DB	73	---FQDFHVSVETPDACGATPATQADVFT	IGYPLVQHAFDGFN	SCFLAYG	QTSGGK	128																								
QY	115	SYSMG-----YKREHGVIPRICODMFRIN	LQDKNLTCTVEVSY	LEINERVDLL	168																									
DB	129	TYTMGADVLSALSGGNGVTPICLIEIFARK	SAVEAQGHSRWIV	ELGYVEVY	NERVSDLL	188																								
QY	169	NPSTKG-----NLKVRHPSSTGYPVED	LAKLVYRSFQET	ENLMDGK	KARTVAATNM	220																								
DB	189	GKRKGVKGGGEVYDVREHPSRGVFL	EGQRLVEVGLDDV	VRLIEIGN	GVHRTASTNM	248																								
QY	221	NETSSRSHAVFTLT---	TQKHDEETKMDTEK	VAKISLVLD	LAGSERAT	STGATGARKL	276																							
DB	249	NDRSSRSHAIIMLLREERTMTTKSET	IRTAGKSSRN	LVLAGSER	VAQSOVEG	QOQFK	308																							
QY	277	EGAEINRSLS	TLGRVTAAL	ADMS--GK	KNQLVYRDS	VLTLLKDS	LGGSMTAMIAA	335																						
DB	309	EATHINLS	TLTGRVTDV	LADMATKAKA	QYSVAPFR	DSKLTFL	ILKDSLGGSMTFM	368																						
QY	336	ISPADINFEETLST	LYADSAKRI	NHVVNED	PNARMIREL	KEELAQLR	SKLQSSGGG	395																						
DB	369	VSPSALNYEETLST	LYASRARDI	VVAQVNE	DPARRR	RELEEQ	EDM	RQAM-----	421																					
QY	396	GGAGSGG	GPVEE-----	STPDP	TPLEK	417																								
DB	422	--AGDPAYVSEL	KKLALLEBAQ	KRAADLO	ALERERE	HNQVQER	LLRAT	EAKESELES	479																					
QY	418	QIVSIQ	-----	PDATV	KKMSKAE	IVEQL	441																							
DB	480	RAAALQ	EMTATRRQAD	KMQALN	RLKEQ	ARKERELL	KEMAK	DAALSKVRR	KDAEIA	539																				
QY	442	NOSEKLYDL	NOTWEK	LAKTEIEI	HKERE	AAELG	ELISIE	KGVG	PHYH	KEMPHLVN	LS	501																		
DB	540	SEREKL-----	ESTVAQLERE	QERE	VALDAL-----	QTHOK	KLOE	ALESSE	581																					
QY	502	-----DPLLA	ECLVYNIK	PGQTRV	GNVQD	TQAE	FLNGSK	ILKE	HCIF	ENV	DNVTIV	555																		
DB	582	RTAAERD	OLLQ--LT	ELQ	SERTQL	SOVVD	RRLTR-D	IQRIQ	YE	GETEL	ARD	VALCA	638																	
QY	556	PNE-----	KAAMVNG	VRIDK	PTLRSGY	RIILG	DFHIF	FRN	HP	EEA	ERQ	BSLLRHS	610																	
DB	639	AQEMEARY	HAAPV	PHLOT	LEL	ATE	WEDALR-----	ERALA	ERDE	AAAAA	--	682																		
QY	611	VTNSQL	GSPAPGR	HDRTL	SKAGSD	ADGDS	SDSPL	PHFR	KDS	DNF	YAR	REA	SAI	LGLD	670															
DB	683	-----	LDAAAST	SQN-----	ARS	ACER	LTSLE	705																						
QY	671	QKISHL	TDELD	ALDFD	VDQKAR	AVRR	GLVED	NE	DS	SQSS	FP	VP	RV	KY	NS	NGT	IND	SLDT	730											
DB	706	QOLRE-SE	ERAAEL	ASQ	EATAA	AKSSAE	QDRE	NT	RAT	L	EQ	L	RES-----	E	ARAAEL	757														
QY	731	AITMPG	TPSR	DDGD	ALFF	GDKK	SQD	ASNV-----	D	VEL	ROO	Q	Q	A	ME	EAL	TAK	781												
DB	758	ASQ	EATAA-----	KMSAE	QDRE	NT	RAT	L	EQ	L	R	D	S	E	E	A	A	E	A	L	S	Q	E	S	T	T	A	A	K	807

RESULT 7

us-08-282-845-2

; Sequence 2, Application US/08282845

; Patent No. 5719263

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

```

Query Match      20.2%; Score 814; DB 1; Length 955;
Best Local Similarity 28.1%; Pred. No. 6.9e-60;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY      7  IKVVRVVRPNARE--IDRGAKCTVRMEGNOTILTPP-----PGAEKARKSGKTTMD 57
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      13  VKYSVRVRPLNERENNAPEGTKVTVAAKQAQAAVVTVKVLGSSNNSGAESMGTARRVAQD 72

QY      58  GKAPAFDRSYWSP---DKNAPYARQEDLFDLGVPLLDNAFKNQYNCICFAYGQTGSGK 114
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      73  ----FQPDHFVWSVETPDACGATPATQADVFTIGYPLVQHAFDFGNSCLFPAYGQTGSGK 128

QY      115  SYSMKG-----YCKEHCYVPRICODMFRRLNELQKDKNLCTVEVSYLEYIYNERNVDLL 168
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      129  TYTMGADVSAISGEGGVTPRICLEIFARKASVEAQGHSRMIVELGVVEYVYNERVDLL 188

QY      169  NPSTKG-----NLKVRHPSTGPVVEDLAKLWRSFQEIENLMDGNKARTVAATNM 220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      189  GKRRKGVKGGEVYVDVREHPSGVFLEGORLVEVGSDDVRLIEIGNGVRHTASTKM 248

QY      221  NETSSRHAVFTLTL----TQKWHDEETKMDTEKVAKISLDVLGASERSTATGATGARLK 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      249  NDRSSRSHAIMLLREERTMTTKSETIRTAKSSRMNLVDLAGSERVAQSGVEGQOFK 308

QY      277  EGAEINSLSTGLCBVIAALADMS--GKQKNQLVPYRDSVLTLLKSLGNSMTAMIAA 335
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      309  EATHINLSLTGLRVIDVLDMATKGAQAQSVAPFRDSKLTFTLKDSLGNSKTFMTAT 368

QY      336  ISPADINFEETLSLRYADSAKRINKHIAVNEDPNARMIRELKELAQRLKSLGSSGGG 395
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      369  VSPALANYEETLSLRYASRARDIVNVAQVNEDPARRIRELEQMEDMROAM----- 421

QY      396  GGAGSGGPGVEE-----SYPPDTPLEK 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Query Match      20.2%; Score 814; DB 5; Length 955;
Best Local Similarity 28.1%; Pred. No. 6.9e-60;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY 7 IKVVVRFPFNARE--IDRGAKCIVRMEGNOITLPP-----PGAEEKARKSKTMD 57
DB 13 VKVSVVRFLNERENNAPEGTKVTVAAKQAAAVTVKVLGGSNNGAAESMGTARRVAQD 72

QY 58 GKAFADFSYNSF---DKNAPNYARQEDLFQDLGVPLLDNAFKGYNCCIFAYGOTGSK 114
DB 73 ----FOFDHVFWSVEPDPACGATPATQADVFTIGYPLVQHAFDGFNSCLFAYGOTGSK 128

QY 115 SYSMMG-----YGKEHGVIPIRCQDMFRINELQDKNLTCTVEYSYLEIYNERVDLL 168
DB 129 TYTMMGADYSALSGEGNGVTPRICLFIKAKASVEAOGHRSWIVELGYVEVYNERVSDLL 188

QY 169 NPSTKG-----NLKVRHPSTGYPVEDLAKLVVRSFOEITENLMDGNKARTVAATNM 220
DB 189 GKRRKGKGGGEEVYVDVREHPSRGVFLQGRVLEQSLDDVVRLLIEINGVRHTASTKM 248

QY 221 NETSSRSHAVFTLTL---TQKHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLK 276
DB 249 NDRSSRSHAIIMLLREERTMTTKSGETIRTACKSRMNLVDLAGSERVAQSQVEGQFK 308

QY 277 EGAEINRSLTGLRVTAALADMS--GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAA 335
DB 309 EATHINLSLTGLRVLDVLMATKGAQYSVAPPRDSKLTFLKDSLGGNSKTFMIAT 368

QY 336 ISPADINFEETSLTRYADSAKIKHVVNEDPNARMIRELKEELAQLRSKLOSGGGG 395
DB 369 VSPSALNYEETSLTRYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAM----- 421

QY 396 GGAGSGGPGVEE-----SYPPDTPLEK 417
DB 422 --AGGDPAYVSELKKLALLESEAKRAADLQALERERHNOYQERLLRATEAEKSELES 479

QY 418 QIVSIQO-----PDATVKKMSAEIVEQL 441
DB 480 RAAALQEBTATRRQADKMQALNLRKEEQARKERELLKEMAKKDAALSKVRRKDAEIA 539

QY 442 NQSEKLYDLNQTWEELKATEETHKREAALELGISIEKFGVPVHSKEMPHLVNLS 501
DB 540 SEREKL-----ESTVAQLERQREVALDAL-----QTHQRKLQEALESSE 581

QY 502 ----DPLLAECILYNIKPGQTVGNVNDQTAEBIRLNGSKILKEHCTETENDVNVTV 555
DB 582 RTAAERDQLQO--LTELQSBRTQLSQVVDRELRTR-DLQRIQYEGYGETELARDVALCA 638

QY 556 PNE-----KAAMVNGVRIDKPTLRSGYRILIGDFHIFRPNHPEEARAQESLLRHS 610
DB 639 AQEMEARYHAAVFFHLOTLELATWEDALR-----ERALAERDEAAAAE-- 682

QY 611 VTNQLGSPAPGRHRTLSKAGSDGDSRSDSPLPHFRGKSDSWFYARREAAASAILGLD 670
DB 683 -----LDAAASTSQN-----ARESACERLTSL 705

QY 671 QKISHLTDELDFDDVQKARVRRGLVEDNEDSDOSSPPVRDKYMSNGTIDNFSLDT 730
DB 706 QQLRE--SEERAELASQLEATAAAKSSAQEDRENTATLEQOLRES-----BARAEL 757

QY 731 AITMPTGTPRSDDGDALEFFGDKSKQDASNV-----DVEELRQQAQMEALKTAK 781
DB 758 ASQLEATAAA-----KMSAQEDRENTATLEQOLRDSERAAELASQLESTTAAK 807

```

```

RESULT 9
US-08-428-414A-3
; Sequence 3, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: LEISHMANIASIS

```

```

; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecck, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-428-414A-3

```

```

Query Match      20.2%; Score 812.5; DB 2; Length 955;
Best Local Similarity 40.0%; Pred. No. 9.2e-60;
Matches 201; Conservative 75; Mismatches 164; Indels 63; Gaps 13;

```

```

QY 7 IKVVVRFPFNARE--IDRGAKCIVRMEGNOITLPP-----PGAEEKARKSKTMD 57
DB 13 VKVSVVRFLNERENNAPEGTKVTVAAKQAAAVTVKVLGGSNNGAAESMGTARRVAQD 72

QY 58 GKAFADFSYNSF---DKNAPNYARQEDLFQDLGVPLLDNAFKGYNCCIFAYGOTGSK 114
DB 73 ----FOFDHVFWSVEPDPACGATPATQADVFTIGYPLVQHAFDGFNSCLFAYGOTGSK 128

QY 115 SYSMMG-----YGKEHGVIPIRCQDMFRINELQDKNLTCTVEYSYLEIYNERVDLL 168
DB 129 TYTMMGADYSALSGEGNGVTPRICLFIKAKASVEAOGHRSWIVELGYVEVYNERVSDLL 188

QY 169 NPSTKG-----NLKVRHPSTGYPVEDLAKLVVRSFOEITENLMDGNKARTVAATNM 220
DB 189 GKRRKGKGGGEEVYVDVREHPSRGVFLQGRVLEQSLDDVVRLLIEINGVRHTASTKM 248

QY 221 NETSSRSHAVFTLTL---TQKHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLK 276
DB 249 NDRSSRSHAIIMLLREERTMTTKSGETIRTACKSRMNLVDLAGSERVAQSQVEGQFK 308

QY 277 EGAEINRSLTGLRVTAALADMS--GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAA 335
DB 309 EATHINLSLTGLRVLDVLMATKGAQYSVAPPRDSKLTFLKDSLGGNSKTFMIAT 368

QY 336 ISPADINFEETSLTRYADSAKIKHVVNEDPNARMIRELKEELAQLRSKLOSGGGG 395
DB 369 VSPSALNYEETSLTRYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAM----- 421

QY 396 GGAGSGGPGVEE-----SYPPDTPLEK 417
DB 422 --AGGDPAYVSELKKLALLESEAKRAADLQALERERHNOYQERLLRATEAEKSELES 479

QY 418 QIVSIQO-----PDATVKKMSAEIVEQL 441
DB 480 RAAALQEBTATRRQADKMQALNLRKEEQARKERELLKEMAKKDAALSKVRRKDAEIA 539

QY 442 NQSEKLYDLNQTWEELKATEETHKREAALELGISIEKFGVPVHSKEMPHLVNLS 501
DB 540 SEREKL-----ESTVAQLERQREVALDAL-----QTHQRKLQEALESSE 581

QY 502 ----DPLLAECILYNIKPGQTVGNVNDQTAEBIRLNGSKILKEHCTETENDVNVTV 555
DB 582 RTAAERDQLQO--LTELQSBRTQLSQVVDRELRTR-DLQRIQYEGYGETELARDVALCA 638

QY 556 PNE-----KAAMVNGVRIDKPTLRSGYRILIGDFHIFRPNHPEEARAQESLLRHS 610
DB 639 AQEMEARYHAAVFFHLOTLELATWEDALR-----ERALAERDEAAAAE-- 682

QY 611 VTNQLGSPAPGRHRTLSKAGSDGDSRSDSPLPHFRGKSDSWFYARREAAASAILGLD 670
DB 683 -----LDAAASTSQN-----ARESACERLTSL 705

QY 671 QKISHLTDELDFDDVQKARVRRGLVEDNEDSDOSSPPVRDKYMSNGTIDNFSLDT 730
DB 706 QQLRE--SEERAELASQLEATAAAKSSAQEDRENTATLEQOLRES-----BARAEL 757

QY 731 AITMPTGTPRSDDGDALEFFGDKSKQDASNV-----DVEELRQQAQMEALKTAK 781
DB 758 ASQLEATAAA-----KMSAQEDRENTATLEQOLRDSERAAELASQLESTTAAK 807

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: H.sapiens
US-09-541-782-10

Query Match      16.7%   Score 673;   DB 4;   Length 1057;
Best Local Similarity 27.1%;   Pred. No. 6.4e-48;
Matches 245;   Conservative 146;   Mismatches 290;   Indels 224;   Gaps 34;

Qy  4  GGNKVVVRPFPNAREIDRGAKCIVRMENQOTILTPPPGAEERKARKSGKTIIDGPKAFAPD 63
Db  16  GKNIQVVVRCRPFNLAEERKASAHISCECD-----PVRKEVSVRTGLADKSSRKTYT 67
Qy  64  FDRSYWSDFNAPYARQEDFODLGVPLLDNAFKGYNCCIFAYGOTGGSGKSYNMG--- 120
Db  68  FDMVFGASTK-----QIDVRSVVCPIDEVINGYCTIFAYGOTGTGKTFTMEGERS 120
Qy  121  ----YGKEH-----GVIPRICQDMFRRINELQKDKNLCTVEYSYLEIYNERVRLDLPNST 172
Db  121  PNEEYTWEEEDPLAGIIPRTLHQIPEKLT---DNGTEFSVKVSLLEYNEELFDLLNPSS 176
Qy  173  --KGNLKVREHP--STGPYVEDLAKLVRSFQBIENLMDGKNKARTVAATNNNETSSRS 228
Db  177  DVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEVYQILEKGAARKTTAATLNNAYSSRS 236
Qy  229  AVFTLTLLQKHDEETKMDTE---KVAKISLVDLAGSERATSTGATGARLKEGAGINRSL 285
Db  237  SVFSVTI---HMKETTIDGELVKIGKLNLDLAGSNIGRSGAVDKRAREAGINOSL 292
Qy  286  STLGRVIAALADMSSGKQKNQVPYRDSVLTLLKDSLGGNSMTAMIAAISPADINFE 345
Db  293  LTGLRVTALVE-----RTPHPVRESKLTIRLQDSLGGRTRTSIATISPASLNLEE 345
Qy  346  TLSTLRYADSARKRNHAVNEDPNAR--MIRELKEELAEQLAKSKLQSSGGGGGAGGSGP 404
Db  346  TLSTLEYAHRKRNILNKPEVNQKLTAKALIREYTEIERLERKDLAAAREKNG----- 397
Qy  405  V---EESY---PPDTPLEKOIVSIQPDATVKKSKAEIVEOLNQSEKLYRD---LNQ 453
Db  398  VIISENFVMSGKLTVOEEQIVEL-----IEKIGAVE--EELNRVTELFMDNKNELDQ 449
Qy  454  TWEELAKTEIEHKEREAALE--ELGISIEKGFVGPYHSKE-----MPHLVNLSDPDLLA 506
Db  450  CKSDLQNKIQLETTQKHQLQETKQLQVKEEYTSALESTEELKHLDAASKLLNTVEETKD 509
Qy  507  ECLVNIKFGQTRVGNVNDTQ-----AEIRLNGSK-----ILKEHCT--PEN 547
Db  510  VSGLSKLCRKKAVDQHNAAEQDIFGKNLNSLFNNMEELIKDGSCKKAMLEVHKTLFGN 569
Qy  548  -----VDNVVTIVPNEKAAMVGVNRIDKPTLRSGYRIILGDFHIFREHNHPEARA 599
Db  570  LLSVSSVSDLYITTVLALGSLTSIPEN---VSTHVSQIFNMIL----- 608
Qy  600  ERQEOSLLRHVSVT-----NSQLGSPAGRHDRTL 628
Db  609  --KEQSLAAESKTVLQELINVLKTLSSLEMLSPVTVVSVILKINSQ-----KHIFKT 660
Qy  629  SKAGSDADGDSRSDSPLPFRGKSDSDWEYA-----RREAASAILGLDQKISHLTDD 679
Db  661  SLTVADKTEDQKK-----RNSDGLFSLICNLNHELQENTICSLVESQKQCNLTED 711
Qy  680  -----ELDALFDVQK--ARAVRRGLVEDNDSQSFSFPVR 714
Db  712  LKTIKQTHSQELCKLMLNLTWERFCALEKCNIQKPLSVQENIQKSKIVNKNFTFH--S 770
Qy  715  DKYMSNGTIDNFSLDTAITMPTGTPRSDDGDALFFGDKSKQDASVNDVEELRQQAQME 774
Db  771  QKFCADS--DGFSQL-----RNFQEGTKLVEESVKHSDKLNGLNLEKISQETQRC 820
Qy  775  EALKT 779

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: H.sapiens
US-09-541-782-10

Query Match      18.5%   Score 747;   DB 4;   Length 357;
Best Local Similarity 46.6%;   Pred. No. 6.4e-55;
Matches 170;   Conservative 54;   Mismatches 119;   Indels 22;   Gaps 8;

Qy  6  NIKVVVRPFPNAREIDRGAKCIVRMENQOTILTPPPGAEERKARKSGKTIIDGPKAFAPD 65
Db  3  NQVAVVRPLSKXXXXXXXXXXXXXXXXXXXXVNDNRDPDGFDSREKVM-----AFGFD 58
Qy  66  RSYNSFDNAPYARQEDFODLGVPLLDNAFKGYNCCIFAYGOTGGSGKSYNMGYGREH 125
Db  59  YCYSVNPEDEQYASQDVVFDLGEVLGSGVAKGYNICLFAFGTGPSTYMLGTPASV 118
Qy  126  GVIPRICQDMFRRINELQKDKNLCTVEYSYLEIYNERVRLDLPNSTKG---NLKVR 179
Db  119  GLTPRICEGLFVR---EKDCASLPSSCRIRKVSFLFIYNERVRLDKSGQKKSYTLVR 174
Qy  180  EHPSTGPTVEDLAKLVRSFQBIENLMDGKNKARTVAATNNNETSSRSRAVFTLTLLTKW 239
Db  175  EHPGPTVQGLSQHVVTNYKQVIQLLEEGIANRITAATHVHEASSRSHAFTIHYTOA- 233
Qy  240  HDEETKMDTEKVKAKISLVDLAGSERATSTGATGARLKEGAEINRSLTLGRVIAALADM 299
Db  234  -ILENLFSEWASKINLVLDLAGSERADPSYCKD--RIAGANINKSLVTGLVISTLAONS 291
Qy  300  ---SGKQKN--QLVPYRDSVLTLLKDSLGGNSMTAMIAAISPADINFEETLSTLYAD 354
Db  292  QVFSSCQSLNSSYIPYRDSVLTLLKDSLGGNSKTIINAVTVSPAHTSYSETMTSLRYAS 351
Qy  355  SAKRI 359
Db  352  SAKNI 356

RESULT 11
US-09-541-782-10
; Sequence 10, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE OF INVENTION: 1015
; CURRENT APPLICATION NUMBER: US/09/541.782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
```



```

; : | | | | | | | | | | : : | | | | | | | | | | | |
Db 354 LNFQRAKLINKA VNNEDTQGNVSQLQAEVKRLKEQLAEL-----ASQTP 400
QY 404 PVEESY 409
; : | | :
Db 401 PLEHHH 406

RESULT 14
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: A. nidulans
US-09-541-782-2

Query Match 15.7%; Score 633; DB 4; Length 1184;
Best Local Similarity 26.7%; Pred. No. 1.8e-44;
Matches 234; Conservative 161; Mismatches 315; Indels 168; Gaps 35;

QY 6 NIKVVVRFPFNAREIDRGAKCIVRMENQOTILTPPPGAEKARKSKRTI--MDGPKAFA 63
; : | | | | | | | | | | : : | | | | | | | | | | | |
Db 81 SIHVVRRCGRNREVKENSQVVLQTEG-----VKGTVELSMGNPAVS 124

QY 64 FDRSYWFDKNAPYARQEDLFQDLGVPLLDNAFKGNINCIFAYGQSGKSYMMG--- 120
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 125 -NKTY-TFKRVFSAADQITVYEDVPLVITEMLAGYNCTIFAYGOTGTGYTMSGDMT 182
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 121 -----YKHEGVIPRICODMFRRLNEQKDKNLCTVEVSLEYINERVRDLNLPSTGN 175
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 183 DTGLILDNAGIIPRVLSFLAKLADYES-----TVKSPFIYNEELRDLLSAENPK 236
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 176 LKV-----REHPTGPGYVEDLAKLVRSFQEIENLMDGNKARTVAATNMNETSSRSHA 229
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 237 LKIYDNEQKKGHMST--LVQGMETIYDSATAGIKLLQOGSHKQVAATKCNLDLSRSHT 294
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 230 VFTLTLTQKHDEETKMDTEKV--AKISLVDLAGSERATSTGATGARLKEGAELNRSLS 287
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 295 VFTITVNIK-----RTTESGEYVCPGKLNLDVLAGSENIGRSGAENKRATEAGLINKSLT 351
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 288 LGRVIAALADMSSCKKNOLVPYRDSVLTWLLKDSLGGNSMTAMIAISPADINFEETL 347
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 352 LGRVINALVD-----KSHQIPYRESKLRLLQDSLGGRTKTCIIATWSPARNLEETI 404
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 348 STLRYASAKRIKHVYVNEED--PNARMIRELKEALQRLSKLOSSGGGGGAGGGGPVE 406
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 405 STLDYAFRAKNIRKNKPQINSTMPKMTLLREPTAEIKLAEKLIATIRHNGVYMS-----V 459
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 407 ESYPPDTPLE---KQIVSIQOPDATVVKMSKA---EIVEQLNQSEKLYRDLNQTWEKLA 460
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 460 ESY-EEMKMENESRIIIEEO--RAKIESMESLRRHKVQELTLTTSK-FNDLKKNDODTLA 516
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 461 ---KTEERIKREAAELGTSIEKGFGVPYHSEKMEPLHVLNLSDDPLLAELCVYNIKPGQ 517
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 517 ALCTNDVLOQTDIVLQNTRAQLBEEMLCRAHEETER--QLQD---VGKGLISTL--GQ 569
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 518 TRVGNVQ-----DTQAEIRLNGSKILKEHCTFFENVDNVTIVPNEKAAMVNGVRIDKP 572
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 570 T-VEDINSLSQSKLRKAELDATNAELWRASST--EVSQVTKRIDQRYEAFQTRHAKLLET 626
; : : | | | | | | | | | | : : | | | | | | | | | | | |
```

```

QY 573 TRLSGYRIILGDFHIFRNHPPEARERQORQSLRHVSVTNSQLGSPAPGRHD----- 625
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 627 TSVK-----VNEFIATEISNIERTSDLSYENRSLDAACNNAKAETSSAHEDMNVLEE 680
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 626 -----RTLKAGSDADGDSRSDPLPHFRGKSDWDFARREASAILGLDQK-----I 673
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 681 IKDLREEVKSQVGEGLNGLSAAARISEEVIGETQLHSLHTSFNNLCKDLKLSIFETMA 740
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 674 SHLTD--DELDALFDDVOK-----ARAVR-----RGL 698
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 741 THLSEQKNEINRLRAELQSSNRQNIETHKASAHLAQAEIEEHVAEAEREILMSQIKAL 800
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 699 VEDNEDSDSQSSFPVRDKNYKNSGTIDNFSLDLTALTMTGCTPRSDDDGDALFGDKSKQDA 758
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 801 VEESR---QKQF-ARLRAKIDGVRTAISASGDMLEQATTOHROIDEWVF---KSEQFA 852
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 759 SNV-----DVEELRQQAQAMEEALKTAQOE 783
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 853 KDVNASKDEIRTKLQNDWEAFDQNRSTIRKATESVHKE 890
; : : | | | | | | | | | | : : | | | | | | | | | | | |

RESULT 15
US-08-713-815A-4
; Sequence 4, Application US/08713815A
; Patent No. 5830659
; GENERAL INFORMATION:
; APPLICANT: Russell J. Stewart
; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED
; TITLE OF INVENTION: SEPARATIONS BY KINESINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Thorpe, No. 5830659th & Western, L.L.P.
; STREET: 9035 South 700 East, Suite 200
; CITY: Sandy
; STATE: Utah
; COUNTRY: USA
; ZIP: 84070
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: AST Ascentia 900N
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Word Perfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,815A
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Alan J. Howarth
; REGISTRATION NUMBER: 36,553
; REFERENCE/DOCKET NUMBER: T3214/U-2202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (801)566-6633
; TELEFAX: (801)566-0750
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
US-08-713-815A-4
```

Query Match 15.5%; Score 625; DB 2; Length 411;  
Best Local Similarity 39.4%; Pred. No. 1.6e-44;  
Matches 154; Conservative 67; Mismatches 128; Indels 42; Gaps 11;

```

QY 6 NIKVVVRFPFNAREIDRGAKCIVRMENQOTILTPPPGAEKARKSKGTINDGPKAFAD 65
; : | | | | | | | | | | : | | | | | | | | | | | |
Db 12 SIKVFRPRUNDSEERAGSRFVVRP-----PNVNEC-----ISIAKGVYLF 56
; : : | | | | | | | | | | : : | | | | | | | | | | | |
QY 66 RSYWFDKNAPYARQEDLFQDLGVPLLDNAFKGNINCIFAYGOTGSGKSYSM---MGY 122
; : : | | | | | | | | | | : : | | | | | | | | | | | |
Db 57 KVF-----KPN-ASQEKVYNEAAKSIIVTDVLACVNGTIFAYGOTSSGKTHTMESG 109
; : : | | | | | | | | | | : : | | | | | | | | | | | |
```







[illegible]

RESULT 2  
A56921  
Kinesin family protein KIF1a - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 02-Feb-2001  
C:Accession: A56921  
R:Okada, Y.; Yamazaki, H.; Sekine-Alizawa, Y.; Hirokawa, N.  
Cell 81, 769-780, 1995  
A:Title: The neuron-specific kinesin superfamily protein KIF1a is a unique monomeric motor  
A:Reference number: A56921; MUID: 95292344  
A:Accession: A56921  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1695 <RES>  
A:Cross-references: GB:D29951; NID:g976234; PIDN:BAAO6221.1; PID:g976235  
A:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; plectostrin  
C:Keywords: nucleotide binding; P-loop  
E:6-360/Domain: kinesin motor domain homology <MOT>  
E:97-104/Region: nucleotide-binding motif A (P-loop)

Query Match	41.3%	Score 1663.5	DB 2	Length 1695
Best Local Similarity	46.5%	Pred. No. 7.8e-82		
Matches 350	Conservative 126	Mismatches 167	Indels 109	Gaps 15

```
0Y      4 GGNKVVVRPPFNAREIDRAKIIVMEGQITLTTPPGAEEAKRSKGTINDGPAAFA 63
        | : : : | | | | | | : : : | : : : | : : : | : : : | : : : |
Db     3 GASVAVAVRPFNRSREMSRDSCIIOMSGSTTIVVPKOPET-----PSPFS 51
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
0Y      64 FDRSYWSPDKAP---NYARREDLFQDLGYPLDINAKGYNCCIIFAGTGSGKSYSMMG 120
        ||| |||| : : | || : : : : | : : : | : : : | : : : | : : : |
Db     52 FDYSYWS--HTSPEDINVASQKYRRDIGERBMLQHAEFGYNCVFAYGGAGASYTMMG 109
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
0Y      121 YGR--EHGVIPRIQDMFRRIRNELQKNLCTEYVSLEYIENRVADLLNPSTKGLKV 178
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db     110 KQEDDOOGIIPQLCEDLEFSRIDNTTD--NMSYSEVSYMEIGCYERVDDLNPXKGNLRV 168
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
0Y      179 RHHSTGTYVEDLAKLVYRSQPELENMDGNGNRKRYAALNNMETSRSRAVFTTLITOK 238
        |||| | | | | | | : : | : : : | : : | : : | : : | : : | : : |
```

D	b	169	REHPLIGLYVEDLSKLAATVSYNDIJDLMDSGNKPRVTAATNMNTESSRSHAVNIIPTOK	228
Q	y	239	WHDEETKMDTEKVKAKISJYDLAGESEATSGATAPARKKEGAETIRSLSTIGRYTAMADM	298
D	b	229	RHDEETNITTEKYSISVYDLAGESEADSTGAKTIRKEGANINIKSLTTGKYSISALAE	288
Q	y	299	SSG-----KOKKNOLVYRDSVLTWMLKDSLGGNSMTAMIAISPADINEETLSTLRYA	353
D	b	289	DSGPNKKKKKKTDEIPYRDSVLTWMLRENNGNSPTAMVAALSPADINVEDTLSTLRYA	348
Q	y	354	DSARIKNHAVNNDPNARMKRELKEELAOURLKSTOSSGGGGGAGSGGPGVEBESYPTD	413
D	b	349	DRAQOIRNALLIINDPNKKLTRELKDEVTYRLDLYAOQJG-----DIT	392
Q	y	414	PLEKOIVSIOOPDATYVKK-----KAEIYOLNQSEKLYRDLNQTW	455
D	b	393	DMTNALVGM-SPSSSLATSSRAVSGLHIERILFAPGSEATIRLKEETEKIIAELNETW	451
Q	y	456	EEKLAKTEIHKKEKAALEELGISIER--GPGYGHSEKEMPHLYNSDDPLAECLYNT	513
D	b	452	EEKIRFEATIMEERALLAEWGMARDGGTLGVFSPKRPHLYNLNDEPLMSCELLYI	511
Q	y	514	KPGOTRGVNVQDPOAETIRLNGSKILKEHCFEEN-----VDNVYTVIPNKEAMVANGVR	568
D	b	512	KDGYTRAGREDAERROPIVLSGHFIKEBHCFRSDSGGGAVALTEPCGADITYVNGK	571
Q	y	569	IDKTYRLRSGRITILGDFHTRFNMHPEARAEKOEOSILKRHSVYNSOLGSPAPGRHRTL	628
D	b	572	VTEPSILRSGRNRIITGMSHYFRFMHPQARERER-----	606
Q	y	629	SKASDADGDSRSDSPRPHFRGKSDMFYAREREAASAALIGLDOKISHLTGDELDALFDV	688
D	b	607	-----TPCAETPAEYVMAFNQRELLER-qGIDMK-qEMQROIRQELDDY	649
Q	y	689	OKAAVNRGLVEDNEDDGSOSFFRYDKYSN 720	
D	b	650	RREREAEYILLE-qOQRIDYESKLEALOKOMDS 680	

```

RESULT      3
JN0114
K:Kinesin-related protein unc-104 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: JN0114
C:Otsuka, A.T.; Jeyarajakashi, A.; Garcia-Anoveros, J.; Tang, L.Z.; Flisk, G.; Hartshorne
Neuron 6, 113-122, 1991
A:Title: The C. elegans unc-104 gene encodes a putative kinesin heavy chain-like protein
A:Reference number: JN0114; MUID:91097805
A:Accession: JN0114
A:Molecule type: mRNA
A:Residues: 1-1584 <OTS>
A:Cross-references: GB:M58582
A:Note: 598-Thr and 930-Met were also found
C:Genetics:
A:Gene: unc-104
C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; plects
C:Keywords: ATP; microtubule binding; nucleotide binding; P-loop
F:4-353/Domain: kinesin motor domain homology <KMT0>
F:93-100/Region: nucleotide-binding motif A (P-loop)
F:1285-1287/Region: cell attachment (R-G-D) motif
F:99/Binding site: ATP (Lys) status predicted

```

Query Match	39.5%;	Score 1593.5;	DB 1;	Length 1584;
Best Local Similarity	44.7%;	Pred. No. 4.3e-78;		
Matches 356;	Conservative 139;	Mismatches 213;	Indels 89;	Gaps 19;

```
QY      6  NIKYVVRVRPPENAEIDRGAKCIVMEGNOTILTPPGAEEKARKSGKTIIMDGKPAFAD 65
      ::||| ||||| ||| :||:: ||| :|| :||
Db      3  SVKAAVVRVRPPENOREISNTSKCVLQVNGNTTTI-----NGHSINKENEFSENF 50

QY      66  RSYVSEEDKANPNVARKQEDLPDGLGVLPLLDNAFKCYNNCIPAYGOTGSGKSYKSMGTYCK-- 123
```



A: Molecule type: mRNA  
A: Residues: 1-1921 <LIH>  
A: Cross-references: EMBL:U01788; NID: g1906595; PID: g1906596; PIDN: AAB50404.1  
C: Genetics:  
A: Cross-references: FlyBase: FBgn0019968

Query Match 34.7%; Score 1396.5; DB 2; Length 1921;  
Best Local Similarity 48.8%; Pred. No. 2.5e-67;  
Matches 305; Conservative 88; Mismatches 171; Indels 61; Gaps 8;

```
QY 7 IKVAVRPPFARFARDIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPPAFADR 66
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6 IKAVARPPFARRRRIEDDTKCIYMERQQTLLQNPPEKTERKQ-----PPTFAFDH 58

QY 67 SYMSFDKNAAPYARQEDLFODLGYPLLDNAFKGYNNCIFAYGOTSGSKSYMKGKEG 126
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 59 CFYSLNDEDEFASQEVYFPCDVGRIIDNMFAGYNACIFAGQSGSKSYMKTQESKG 118

QY 127 VIPRICODMERRIWELOKKNLCTVEVSYLEYNERVRLDLPNS-TKGNLKVRHPSGTG 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 119 IIPRLCQOLFSAIIN-KSTPELMATKVEVSYWEIYNEKHYHDLDPKPNKQSLKVRHHVMG 177

QY 186 PYVEDLAKIVRSFOELTENLMDENKARTYAATMNETSSSHAVFTLLTQKHDEETK 245
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 178 PYDGLSLQAVTSYQDIDNLTBEGKSTRYAATMMNMAESSSHAVFSVLQILLDTATG 237

QY 246 MDTEKVAKISLVLDAGSERATSTGATGARLKEGAENRSITLGRVIAALADSSGKOK 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 238 VSGEKVSRMSISVLDAGSERAVKTCAGVDRLKEGSKNINSLTTLGLVLSKLADQNGKSG 297

QY 306 N-QLVPRDSVLTWMLDLSGNSMTAMIAISPADINFEETSLTYADSAKRIKHAHV 364
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 NDKVPRDSVLTWMLDNGNSRTYVAVATISPSADNTEETSLTYADSAKRIKHAHV 357

QY 365 VNEDPNARMIRELKEELAQRSKLOSOGGAGGSGGAPVEESYPPDTPLEKQIVSIQ 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 358 VNEDPNARMIRELKHEVETLSMLKHA-----TGSFY----- 389

QY 425 PDATVKKMSAEIYEQUNOSEKLYRDLNQWEKLAETELHKEREALBELGISIEKGF 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 390 -----GDVQDKLAEBENLMKQISQIWEKLVKTERIQNRQALKEKMGISVQAS- 438

QY 485 VGPYHSEKEMHIVNLSDPLLAELGVYNIKPGQTRGVGNVMDQAEIRLNGSKILKEHCT 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 -GKVEKNKXYLVNLDPSELNELLVYLKDRILIGRTISGQOPDIOISGLGIQPEHCY 497

QY 545 FEWVDNVVTIVPEKAAVWNGVRIKPTRLRSGYRIILGDFHIFRNHPHPE----- 595
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 498 ITTEDSGLVMEPVQAGCFVNGSAVAEKTPLQNGDRILKGNHNFRRVNSPRKSNNTSMCAS 557

QY 596 --EARRERQOSLIRHSVINSQSLG 618
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 558 EPQTPAQLIDYINFARDEIMONELSN 582
```

RESULT 6  
T29237  
hypothetical protein F56E3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T29237  
R:Gating S.: Wu, X.  
Submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid F56E3.  
A:Reference number: Z20592  
A:Accession: T29237  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1576 <GAT>  
A:Cross-references: EMBL:U41536; PIDN: AAB52613.1; GSPDB: GN00028; CESP: F56E3.3  
A:Experimental source: strain Bristol N2; clone F56E3  
C:Genetics:

A: Gene: CESP: F56E3.3  
A: Map position: x  
A: Introns: 23/1; 47/2; 125/3; 162/2; 191/3; 274/2; 296/3; 498/3; 532/1; 651/3; 723/2;

Query Match 31.9%; Score 1284.5; DB 2; Length 1576;  
Best Local Similarity 38.2%; Pred. No. 2.1e-61;  
Matches 323; Conservative 120; Mismatches 217; Indels 185; Gaps 23;

```
QY 7 IKVAVRPPFARFARDIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPPAFADR 66
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10 KVAVIRAPPFKRRIEDLTKTSVRIQEQVLAHP--IEEK-----NSKTFEFDH 57

QY 67 SYMSFDKNAAPYARQEDLFODLGYPLLDNAFKGYNNCIFAYGOTSGSKSYMKGKEG 126
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 58 SFGCTDHSYVFAQEVYSHLGSQVYENAFSGYNACIFAGQSGSKSYMKTQPPQPG 117

QY 127 VIPRICODMERRIWELOKKNLCTVEVSYLEYNERVRLDLPNSKGN-LKVRHPSGTG 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 118 IIPVCHNDIFRIQE-TSNSLSFKVEVSYWEIYNERVRLDLPKSSKALKVRHKILG 176

QY 186 PYVEDLAKIVRSFOELTENLMDENKARTYAATMNETSSSHAVFTLLTQKHDEETK 245
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 177 PYDGLSLVANSFQISNLBEGKSTRYAATMMNMAESSSHAVFSVLVYQTLHLLENG 236

QY 246 MDTEKVAKISLVLDAGSERATSTGATGARLKEGAEN----- 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 237 FSGEKVAKISLVLDAGSERAGKTGAVGRLKEGGINKNLVSIPLRNDLEKIDFKESAD 296

QY 283 -----RSITLGRVIAALADSSGKOKNOQVPRDSVLTWMLDLSGNSMTA 331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 297 VYVPRINOKLDFRSLTTLGMYISALAEARN--KKDKPIPRDSVLTWMLDLSGNSRTV 353

QY 332 MIAISPADINFEETSLTYADSAKRIKHAHVNEPNARMIRELKEELAQRSKQSS 391
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 354 MIALISPADINFEETSLTYADSAKRIKHAHVNEPNARMIRELKEELAEVETL----- 407

QY 392 GCGGGAGGSGGAPVEESYPPDTPLEKQIVSIQPDATVKKMSAEIYEQUNOSEKLYRDL 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 408 -----MQITQKKHAELE-----ELERLAESRRLVQNM 437

QY 452 NQWEKLAETELHKEREALBELGISIEKGFVGPYHSEKEMHIVNLSDPLLAELCVY 511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 438 NKSWEERLKEETDLNKRQDLTEIGISIESS--GKVEKDRFVLYMNAADPSINELLVY 495

QY 512 NIKPGQTRGVN-----VNOPD-----QAEIRLNGSKILKEHC--TFEN 547
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 496 YIN-GSAITIGNSELETSRDGLSMTCSDSSRRDDKERTSYLRGLGIMRRHAKMTVEE 554

QY 548 VDNVYT--IVP-NEKAAVWNGVRIKPTRLRSGYRIILGDFHIFRNHPHPEARREROQ 604
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 555 YGGRRLRFLVAPMSSECRICVNGKQITERLLRNGNRLVGMNHFYKNCP--KYMDMEQ 611

QY 605 SLRHSVTSNQLGSPAGCRHDLTSKAGSDAGDSRDSFLPHFRKDSQWTFARREAS 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 612 SIMEDSTMEDY-----NDAMEHVNDANPI-----SS 637

QY 665 AILGLDQKISHLRDELDALFDVQKARAVRGVLVEDNEDSDS--QSSFPVRDKYMSNGT 722
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 638 AV--DQYMSVY-----LKQDEKKAALBOQYAFEFYIOSLT 673

QY 723 IDNFSLDTAITMG-----TPRSSDDGDLFGGKKSKODASNVDEERLQOQAQMERALK 778
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 674 AGGFTPTSTPMT-PGFLPRTITPTGIPPPPPANPKQSVKSFYWAQKEEMFAESLK 722

QY 779 TAKQE 783
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 733 RLKAD 737
```

RESULT 7  
T16759  
hypothetical protein R144.1 - Caenorhabditis elegans



Db 750 DHPVDFOLAHQGIILQKQEQQLARSELEAKRKALITKIEGRQNHADPFERRQCLELEQPK 809

QY 650 GK-USDMFYARREASAILGIDQKISHLTDELDALFDVQKA---RAVRGLVEDNDS 705

Db 810 KYCNSMEMETERQA-----LALAQOOEHPTLRHEDVAVSPAPKSTILEDIORIMLPSEES 865

QY 706 DSQSSFPVAD-----KYSNGTIDNFSL 728

Db 866 LHKTLQMLKKAQTQRCQDLDPLEFROTQTPDEFG 900

RESULT 9

T13796

kinesin-related protein 38B - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: T13796

R:Motina, I.; Baars, S.; Hales, K.; Fuller, M.T.; Ripoll, P.

J. Cell Biol. 139, 1361-1371, 1997

A:Title: A chromatin associated kinesin-related protein required for normal mitotic chn

A:Reference number: 217759; MUID:98060833

A:Accession: T13796

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1121 <MOL>

A:Cross-references: EMBL:Y15247; NID:g2578010; PIDN:CAA75531.1; PID:g2578011

C:Genetics:

A:Gene: tlo

A:Cross-references: FlyBase:FBgn0004374

C:Superfamily: kinesin-related protein Eg5; kinesin motor domain homology

Query Match 25.7%; Score 1034; DB 2; Length 1121;

Best Local Similarity 33.8%; Pred. No. 4.6e-48;

Matches 274; Conservative 130; Mismatches 290; Indels 116; Gaps 21;

QY 6 NIKVVRPRPFARREIDG-AKCIYRMGNGCTILTPPGAEEKAKSKGKTMDGPKRAF 64

Db 120 NMTVAVRPRPLALCTKQVTVNVQVHGNSNELTVQAGSSADA-SAGVT---HFFSY 173

QY 65 DRSYSPFKNAPNVAIROEDLGVPLLDNAFKYNNCFAYAGTGSGKSYSMGY--- 121

Db 174 DQVYSCDPERKNFCAQAKVFEGTARPLIDTAFEGYNNCLFAYGGTGSGKSYSMGIAL 233

QY 122 -----GKEH---GVIPRIQDMFRRIEQLKDNLTCTYEVSTLEYNEVRDLN-- 169

Db 234 DDALDGGPHYEAQIIPECFELFRIEYAKSQOOLVVEVEYFTEIYNEKIHDLISVQ 293

QY 170 -----PSTKGLKVRHPSSTGPYVEDLAKLVVRSFOETENLMDGNGAR 213

Db 294 HAAATGEGSTPQQOQQOQQORPALKTKREHIREGPYVVDLSAHSVDSYSALRWMLAVGNSQR 353

QY 214 TVAATNNETSSRSNAVF--TLLTQKMHDEETKQTE-----KVAKISLVLDAG 261

Db 354 AVASGAMNDKSSRSISFIVTLNLTDLSDDDLSDTOSTASLRQTRRSKISLVLDAG 413

QY 262 SERATSGTGARLKEGAELNRSLSLTGIVTALADMSS-----GQKKQQLPYPDS 314

Db 414 SFRIVSGSGNGRIRIEGVYSINKSLTLTGKIVTALDSRAKIANGLPGSGTPTFPIRES 473

QY 315 VLTWLLDKSLGNSMTAMIAIASPADINEEFTLSLTARYADSAKRIKNNAVVEDNPAMI 374

Db 474 VLTWLLRENLGNSKTYMLATISPASINADETLATLRACKARSLVNNKYKNVESHDKII 533

QY 375 RLKLELLAQLRKLOS SGGGGGGGAGSGPYVEESTPPDTPLEKQIVSYIQQPDATYKKMSK 434

Db 534 RLRLAEVRLKELRLNEYERQRLSGNSNNPVRKIIIEF-----SYDETEVEALRQOL 586

QY 435 AEIVQLQNSKELTDLNLTWEKLAETLKEIKEREALALEELGISIEGFGVPYHSKMP 494

Db 587 AERERELSAQR-----SMWEKLEKAEADQKSELRLVLRKGLLEL-----TAEQQA 634

QY 495 HLVLNLSDDPLAECLVYNIKPGQTRVG-----NVNODTQAEIRLNGSKILKECHTFENV- 548

[illegible]

Oy	335	SAKRIKHAHVNEEDPNAMRITREKEBLAQRSLQSSGG-----GGGAGSG	402
		:       :       :       :       :	
Db	338	RAKNIKKANAKINEDPKALLREFOKETELKQISSEGGDLDDDESSGEESGDEEAGS	397
Oy	403	G--FVESYPPDPTLEKQVYSIQDPATAYKKM---KAETVEQLNOSKLYINDLQMTWE	457
		:   :   :   :   :   :   :   :   :   :   :	
Db	398	GVKKRRKGKNNRKRLTSPEIMAAQKIDEEKKALEEKKMVEB-----DRMTVIRE	448
Oy	458	KLAETEELHK---BREALEDELGISTEKGYG	486
		:   :     : : :   :   :   :	
Db	449	LQRESELHKAODQOKILTEKLNAIOKLLVG	480

RESULT 11  
A:57107  
Kinesin-related protein KIF3B - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 19-Jan-2001  
C:Accession: A57107  
R:Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.  
J. Cell Biol. 130, 1387-1399, 1995  
A:Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubulin  
A:Reference number: A57107; MUID:96032268  
A:Accession: A57107  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-747 <YAM>  
A:Cross-references: GB:D26077; NID:g1060922; PIDN:BA05070.1; PID:g1060923  
A:Experimental source: brain  
C:Complex: heterodimer with KIF3A (PIR:B44259); the KIF3A/3B heterodimer associates with  
C:Function:  
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal trans  
C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology  
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide  
F:1-363/Domain: head globular #status predicted <HG>  
F:10-346/Domain: kinesin motor domain homology <KMO>  
F:96-103/Region: nucleotide-binding motif A (P-loop)  
F:364-592/Domain: helical rod #status predicted <ROD>  
F:594-747/Domain: tail globular #status predicted <TG>  
#102/Binding site: ATP (Lys) #status predicted

```

Query Match          21.0%: Score 844.5: DB 1: Length 747:
Best Local Similarity 32.9%: Pred. No. 4,5e-38:
Matches 256: Conservative 116: Mismatches 253: Indels 153: Gaps 27:

OY      6 NKKVYVRVPFFARE----IDRGACIYMEGNOQTILTPPGAEKARKSGKTINGCPKA 61
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      9 SVRVVVRCPNPKGKEAKASDYKRVVDVVKL--GQVSVKPKG-----TSHMPKT 56
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY     62 FAFDRSY--MSFEPKNAPNYARQEDFLQDLGLPPLIDNAFKYNNICIFAYGQSGSGSYSMG 120
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     57 FTFDAYVDMN-----AKQFELYDETFRPLVDYSLQGFNGTIFAYGQGTGKITYTMSG 108
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY    121 Y--GKEGVIPRICQDMFRINELQKDKNLTCTVEVSYLETIYNEKRVLDLPSTGKNRK 177
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    109 VAGDEKRGKVIPINSFDHITFTHSRSONQYL--VRASYLETIYQEIRLDLSKDQYKRL 165
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY    178 VREHSTGPIYVDLAKLVYRSFOEIENTLMDEGNKATVAATINNETSSSHAVFTILTRO 237
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    166 LKERPDGTGYVVDLSSTFYKSYKEIEHVANVGNONMSVATNNNEISSSHAIFVTI-- 223
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY    238 KWHDEETKMDPE--KVAKISLVLAGSERATSTGATGTAIRKEGARINSLSTLGVIVA 294
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    224 --ECEVEVIGDENHNRVGLNVLVDLAGSERQAKTGAQGRLEATNTINLSLGLNVISA 281
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY    295 IADMSGKOKKQNLVPIRDSVLTWLTLDKSLDGSNSMTAAITASPADINFEETLSTIRYAD 354
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    282 LVD-----GKSTHPIYRDSKLTLLRLQDSLGSNAKTVMANVANGPASYNVEETLTIRYAN 335
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY    355 SAKRIKNHAVVEDPNARMIRELKEELQALRSKLSSGGG-----GGSGAGSGGP 404
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 336 RANIKKPNVEDPKDALLREPOBELANLKAOLEKSIIGRRRRRREKREGGSGGGEG 355
Qy 405 VE---ESYPPDTPLEKOIVSIQDPATVKKMSKAIVE--QLNOSEKLYLNDLQOTMEKL 459
Db 396 EEECECECEEDGDCKDYWRQOEKLEIKKA---IYEDSHVAEEKM-----RLKEKE 447
Qy 460 AKTEIHKERZALBEIGISTE---KGFGPY---HSEMPILVLSDDPLAECLVY 511
Db 448 KMEDELREKDA--EMIGAKIKAMESKILVGGKNIYDHTNEOOKILEOKROIAEQ--- 5020
Qy 512 NIKPQGTGVGNVDQOAEITLNGSKILKEHCPEUNDVYTYIPNKAAVMVNGVARDK 5711
Db 503 -----KRREIEIQ--QWESDEETLEKE--TTSILOQEVDI-----KTK 5400
Qy 572 PTRLNSGRIILGDFHIFRPNHPEARAEROES-----LIRHSTVNSOLG----- 6171
Db 541 LKLFESKIQAVKAEIHDOEHIERGELBETQNELRLRELKILILENFIPLBKKKIM 6000
Qy 618 -----SPAGRHDRTLSKAGSDAGDSRDS-----PLPHFRGDS 6533
Db 601 NRSFEDDEEDHWKLHPTRLLENQOMMKRPVSAVYKKRPLSOMHAMSMMIRPREYRAEN 6559
Qy 654 DMFYARRAASAILIGLOCKISHLDDDELDAFDVQYKARVARRGLVEDNEDSDSOSAF 711
Db 660 -----IMLELIDPSTTDYDEBPALSPVQ--ALDADALOEDDITQVADSSF 705

```

RESULT 12

S58691

kinesin-related protein KRP95 - sea urchin (*Strongylocentrotus droebachiensis*)

N:Alternate names: kinesin-2 chain B; KRP (85/95) 95k chain

C:Species: *Strongylocentrotus droebachiensis*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001

C:Accession: S58691

R:Rashid, D.J.; Wedeman, K.P.; Scholey, J.M.

J Mol. Biol. 252, 157-162, 1995

A:Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin,

A:Reference number: S58691; MUID:95404610

A:Accession: S58691

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: mRNA

A:Residues: 1-742 <RAS>

C:Complex: heterotrimer of a 115k chain and two kinesin-related chains of 85k (PIR:S58691)

C:Superfamily: kinesin-related protein KRP3; kinesin motor domain homology

C:Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop

F:9-345/Domain: kinesin motor domain homology <KNOT>

F:95-102/Region: nucleotide-binding motif A (P-loop)

F:101/Binding site: ATP (Lys) #status predicted

	Query Match	20.6%	Score 828.5	DB 1	Length 742:	
	Best Local Similarity	31.5%	Pred. No. 3,3e-37,			
	Matches 250;	Conservative 122;	Mismatches 286;	Indels 135;	Gaps 25;	
Qy	7 IKVVVRVREPFNAREIDRGACIVRMEGNOTI-LPPPGAEAKRKSGKTIINDGPKAFNF	64	:	:	:	:
Dd	9 YKVVVRCRPMSNKESIQGHKRIVEMDNKRGLVEVTNPKCPPEGPKNKS-----PFE	58	:	:	:	:
Qy	65 DRSY-WSPDKNAPNPARODLFDODLGVLPLDDNAFGYNNICIFATGYGTGSGRYSMMGYCK	123	:	:	:	:
Dd	59 DRVYVMNN-----SKQIDLDETFRSLVESVDGFNGTITAYAGDTGGTKFTMEGVRS	110	:	:	:	:
Qy	124 E--HGVIPIRLCODMFRIRNELLOKDKNILTCYEVSLEYLEINERVDLLPSPKGMKYKE	180	:	:	:	:
Dd	111 NPELGGVLPNSFEHIFTHTAFTQNOOFL--VRASLEYETGEIRDLLAKDKOKRRDLKE	167	:	:	:	:
Qy	181 HPSSTPYVEDLAKLVVBSFOEILEMDGBGNKARFYAAATMNNTSSRHAVFFLTLDQKKH	240	:	:	:	:
Dd	168 RDTGYVYADLSSFTKSYKTEIEHWMTVGNNNRSGVSTMNHSSSHAIPIFTTT---E	223	:	:	:	:
Qy	241 DEETKMDE--KVAKISLVLDAGSERATSGATGARLRKEGAIEINKSLSLTIGRIVAALAD	297	:	:	:	:
Dd	224 CELSVVDGEGNHRVKKLINVLADLAGSEPOAKTGATGRLKEARKINLSIALNYSALVD	283	:	:	:	:

[illegible]

```

QY 116 KSMNYGGE---HGVTRIQDMPRRINELQKXNLCTVEVSLLEYINRVDLLNPSF 172
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 105 HTMEGKDEPPELRKILIPTEFYRYFEITLARDSGTKEF--LVRSLEYINEVRDLGKD 162
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 173 KGNLKVREHPSTGTYVEEDLALVLRSFQETENIMLDEGNKARTVAATNMNETSRSNAVPT 232
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 SKKMEKLKSPRGVYVDLSQFVCKNYYENMKVLACKDRNQVCATILMNDSSSHSIFL 222
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 233 LFL-----TQKMHDEETKMDTEKVAKISLYDLAGSRATSTATGARLKEGEIN 262
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 223 ITIEIEKLESAAQOKPAKKKDDSNHNVKGLNLVNDLAGSRQDQTKATDRLKEGKIN 282
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 283 RSLSTLGVITVALADMSSGQKOKNQLVPRPSVLTMLLKDSLGSNSMTAMIAISPADIN 342
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 283 LSLTALGVITVALDQKSGH-----IPYRSKLTIRLLQDSLGSNTITVYVANNIGPADWN 336
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 343 FEETLSTLRVADSARKIRKNAHAVNEDPNARMIRELKEBELAOLRSKIQSSGGGGG---A 398
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 337 YDEFTSLTRYANFRANKNIQNKPKINEDPKDAMLROFQEEIRKLKEQLAARAAGCGPITMP 396
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 399 GSGSGPVE-----ESTPPD-----TFLEKQIYSIQQPDATVKKMSAEIVELQ 443
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 397 SGGSGPQKIVERTVEYDPPIDALKAQMRABELEAKMSDSTETLADKARREBAAK-KQ 455
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 444 SEKIVRLNQTWEEKIATKEETIEHKREAEALTELGSIEK 482
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 456 LQAIIIDQKTPAQKKAARDALAKQAEARALAG-AIEK 493
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 14
A47334
Lekin kinesin-related antigen - Leishmania chagasi (fragment)
C.Species: Leishmania chagasi
C.Date: 21-Sep-1993 #sequence_revision 16-Nov-1994 #text_change 02-Feb-2001
C.Accession: A47334
R.Burns Jr., J.M.; Shreffler, W.G.; Benson, D.R.; Chailb, H.W.; Badaro, R.; Reed, S.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 775-779, 1993
A.Title: Molecular characterization of a kinesin-related antigen of Leishmania chagasi
A.Reference number: A47334; MUID:93133867
A.Accession: A47334
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-955 <BDU>
A.Cross-references: GI:107879; NID:9308884; PID:AAA29254.1; PID:9308885
A.Experimental source: MHOM/BR/82/BA-2.C1
A.Note: sequence extracted from NCHI backbone (NCBIT:122864, NCBI:P.122865)
C.Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C.Keywords: ATP; nucleotide binding; P-loop
F.13-398/Domain: kinesin motor domain homology <KMOT>
F.122-129/Region: nucleotide-binding motif A (P-loop)

Query Match 20.1%; Score 814; DB 2; Length 955;
Best Local Similarity 28.1%; Pred. No. 2.8e-36;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY 7 IKVVVVRFPFNARE--IDRGACIVRMENQOTILTP-----PGAEEKAKSGKTIIND 57
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 13 VKASVAVRPLNEENNAPRGTKYTAAKAAAVVIVYKLGSSNGSAEMSGTARVAAD 72
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 58 GPKAFADRSYWF---DKNAPNFAQEDLFQDLGVPLIDNNAFKYINNCIFAYGOTGSGK 114
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 73 ----FQFDHVFMSVETPDACGATPAQADVFTRTIGYPLVQHAFFDFNSCLFAYGOTGSGK 128
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 115 SYSMAG-----YGRKHGIVPRICODMFRRIEQLQKDKNLCTVEVSLLEYINEVRDL 168
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 129 TYTMMGADVSAISGEEGNTVPRICETIEFARKASVEAQGSRNIVELGVIEVNERVSDL 188
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 169 NPSTKG-----NLKVRHPSGTGYVEDLAKLVRSFQETENIMLDEGNKARTVAATNM 220
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 189 GRKKKGKGGGEVYVDVADVNEHPSRGVFLDGQRLVEVGSILDVVRLLIEIGNGRHASTKM 248
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```



QY 221 NETSSRAVFTLT-----TQKHDEETKMDTEKAKISLVNLAGSERATSGATGARLK 276  
 DB 249 NDRSSRSNAITMLLRERTMTTKSGETIRTAGKSSRMNLDVLASERVAQOQVGGQDFK 308  
 QY 277 EGAELNRSISTGRVIAALADMS--GKOKKNQVPRDSVLTWMLKDSGNSMTAMTA 335  
 DB 309 EATHNLSITIGRIDVADLMATGAKAQSVAPFRPSKLTFLIKDSIGSKSPFMAT 368  
 QY 336 ISPADINFEETLSTLRVADSARIKNHAVNEDPNAMRIELKEELADLRSLTSSGGG 395  
 DB 369 VSPSALNTEETLSTLRASRARDIVNAQVNEDPARARIRELEQMEDMRQAM----- 421  
 QY 396 GGAGSGCGPVE-----STPPPTPLEK 417  
 DB 422 --AGDPPAVYSELKKKLLLESEAKRADIALEREREHNOVORLLRATNAEKSELS 479  
 QY 418 QIVSIQ-----PDATVKMKKAEIVEL 441  
 DB 480 RAAALQDEETATROADKQALNLLKEQAKERELLKMAKKDAALSKVRKKDAEYA 539  
 QY 442 NQSEKLYDNLQWTEKLAETEEIHKERPALEELGISIEKGFVGPYHSKNPHLVNSD 501  
 DB 540 SREKLT-----ESTVAQLEREQREVALDAL-----QTHQRKIQEALSESE 581  
 QY 502 -----DPLAECLVYNNKPGQTRVGNVQDTQAEIRLNGSKILKEHCTFFENVAVTIV 555  
 DB 562 RTAERDLDLO--LTELQERTOLSOVATDRELRTR-DLQRIQYEGTEELARVALCA 638  
 QY 556 PNE-----KAAVNVGVNRIDKPTRLRSGRYRIIGDFHIFRPNPEARAERQEOSLRHS 610  
 DB 639 AQEMARVHAANVFIQTLLELATIEMEDALR-----ERALAEERBAAAEE-- 682  
 QY 611 VTNSQLSGPAPGRHDTLSKAGSDADGDSRSDPLPHFRGKDSDFYARREAAASILGD 670  
 DB 683 -----LDAASTSQN-----ARESACERLTSL 705  
 QY 671 OKISHLTDELDALFDYQKAAVRRGLVEDNEDSDSOSFPYRDKYMSNGTIDNFSLD 730  
 DB 706 QQLRE--SEERAAELASQLEATAAKSSAQEDRENTFATLEQQLRES-----EAAAL 757  
 QY 731 AITMPGPRSDDDGALFEFGDKSKQDASNV-----DVEELRQOQAMEEALTKAK 781  
 DB 758 ASQLATAAA-----KMSAQEDRENTFATLEQQLRDBERAAELASQLESTTAAK 807  
  
 RESULT 15  
 B44259  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 19-Jan-2001  
 C:Accession: B44259; S27872  
 R:Alaawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.  
 J. Cell Biol. 119, 1287-1296, 1992  
 A:Title: kinesin family in murine central nervous system.  
 A:Reference number: A44259; MUID:93077686  
 A:Accession: B44259  
 A:Molecule type: mRNA  
 A:Residues: 1-701 <Air>  
 A:Cross-references: EMBL:D12645; NID:g220469; PIDN:BAA02166.1; PID:g220470  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBI:118911)  
 C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with  
 C:Function:  
 A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra  
 C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology  
 C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide  
 F:1-368/Domain: head globular #status predicted <RG1>  
 F:15-351/Domain: kinesin motor domain homology <KMOT>  
 F:100-107/Region: nucleotide-binding motif A (P-loop)  
 F:369-599/Region: helical rod #status predicted <ROD>  
 F:600-701/Domain: tail globular #status predicted <TGL>  
 F:106/Binding site: ATP (lys) #status predicted

Query Match 19.9%; Score 800; DB 1; Length 701;  
 Best Local Similarity 31.8%; Pred. No. 1e-35;  
 Matches 231; Conservative 110; Mismatches 231; Indels 154; Gaps 25;  
  
 QY 6 NIKVYVRVPRFAREID---RGAKCLVRMEGNOTILTPPGAEBAKARKSGKITMOPKAF 62  
 DB 14 NVKVVVRCRPIILEREKSMCYROAVSVDEMRGITV-----HRTDSSN---EPPKTF 61  
 QY 63 AFDRTWSYEDKNAAPYARQEDLFODLGVPLLDNAEFGYNNCIFAYGQTSGRKSYSMG 122  
 DB 62 FEDYVY-----GPE-SKLDVYNLTARPIIDSVLECYGTFTFAIGQITGCTKTFMEGR 114  
 QY 123 KE--HGVIPIQCDMFRRIINELQDKNLTCYEVSYLEYNERVADLLNPSTKGLKVR 179  
 DB 115 AVPGLRGVIPNSFAHIFGHIAKAEGDTR--LYRVSLEYIYNEEDVRLDKGQOTQLEVK 172  
 QY 180 EHPSTGPYEDLAKLVYRSFOELENLMDGKARVYAATNMETSRSRAVFTLT--TQ 237  
 DB 173 ERPDGVYIKDLASAVYVNNADMDRIMTIGHKNRSYGATNMHESSRSHAFITITESE 232  
 QY 238 KMHDEETKMDTEKAKISLVNLAGSERATSGATGARKEGAELNRSISTGRVIAALAD 297  
 DB 233 KGVDSGMHY---RMGRLHVLNLAGSERAKTGATGRLKEATKINLSLTGNSVSLAVD 289  
 QY 298 MSSGKQKNQVPRDSVLTWMLKDSLGNSMTAMIAISPADINFEETLSTLRVADSAR 357  
 DB 290 -----GKSTHVYRNSKLTRLDLSLGNSKTMCMCANIGPADYVDETISTLRVANKAK 343  
 QY 358 RIKNHAVNEDPNAMRIELKEELADLRSLTSSGG--GGGAGSGCGPVE----- 407  
 DB 344 NIKNHARINEDPRDALNLFQKEIEELKRLKEGEVSGSDISGSEDEBELGEGDK 403  
 QY 408 -----SYPPPTPLEKQ-----IYS 421  
 DB 404 RKKRDQACKKVVSPDKMVEQAKIDEERKALETKIDMEERKNRARAELERREKDLIA 463  
 QY 422 IQQPDATVKKMSKAE-----IVEQLNSEKLYRDINQWTEKLAETEEIHKEREA 472  
 DB 464 QOEHOSILEKLSALEKKYIVGGVDLLAKAEQEKLEESNMELERRRRAEQLRKLEEK 523  
 QY 473 -----LEELGISIEKGFVGPYHSKNEMPHLVNSDPLAECLVYNNKPGQTRVGNVND 526  
 DB 524 EGERLDIEEKYVSLDEAAG--KTKRLKRVWTM-----LMAKSEMDLQOE 568  
 QY 527 TOAEIR--LNGSKILKEHCTFFEN--VDNV-----TIVPN-----EKAANVY 564  
 DB 569 HQREILGLEENRQLSRELRDMLIIDNFIPODYQMIENYVHWNNDIGEMOLKCYAVTG 628  
 QY 565 NCVRIDKPT--RLRSGRYRIIGDFHIFRPNHPEARAERQDOSLNRHVTNSQLSGPAP 621  
 DB 629 NNMKQKTPVPDKKERDPEFVDS--HYV-LATYEEL--RQSLMKLERPRTSGKARPAT 683  
 QY 622 GRHDP 627  
 DB 684 GRRKR 689

Search completed: July 25, 2002, 05:46:13  
 Job time: 452 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 25, 2002, 05:39:31 ; Search time 25.37 Seconds

(without alignments)  
1196.536 Million cell updates/sec

Title: us-09-235-416-1

Perfect score: 4030  
Sequence: 1 MSGGNIKVVVRRPFPNARE.....ELRQQAQMEALTKAKQEP 784

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	41.4	1103	1 KFLC_HUMAN	043896 homo sapien
2	1663.5	41.3	1690	1 KFLA_HUMAN	012756 homo sapien
3	1663.5	41.3	1695	1 KFLA_MOUSE	P33173 mus musculu
4	1660.5	41.2	1816	1 KFLB_MOUSE	Q60575 mus musculu
5	1623.5	40.3	1816	1 KFLB_HUMAN	036787 ratu
6	1617	40.1	1097	1 KFLD_RAT	P23678 caenorhabdi
7	1593.5	39.5	1584	1 U104_CAEEL	088658 ratu
8	1554.5	38.6	689	1 KFLB_RAT	P48872 strongyloe
9	860	21.3	699	1 K122_STRPU	015066 homo sapien
10	848.5	21.1	747	1 KFLB_HUMAN	Q61771 mus musculu
11	844.5	21.0	747	1 KFLB_MOUSE	P46871 strongyloe
12	827.5	20.5	742	1 K121_STRPU	P46869 chlamydomon
13	825.5	20.5	786	1 FL10_CHURE	P46865 leishmania
14	814	20.2	955	1 K1NL_LEICH	P26741 mus musculu
15	800	19.9	701	1 KFLA_MOUSE	Q09496 homo sapien
16	789.5	19.6	702	1 KFLA_HUMAN	Q35174 mus musculu
17	782.5	19.4	1231	1 KFLA_MOUSE	Q09239 homo sapien
18	774	19.2	1232	1 KFLA_HUMAN	P43066 mus musculu
19	772	19.2	928	1 K1NH_NEUCR	O51165 ratu
20	770.5	19.1	796	1 KFLC_MOUSE	014782 homo sapien
21	769.5	19.1	796	1 KFLC_RAT	Q09640 gallu
22	765.5	19.0	793	1 KFLC_HUMAN	P46867 drosophila
23	755.5	18.7	1225	1 KFLA_CHICK	Q09178 xenopus lae
24	746.5	18.5	784	1 KFLB_DROME	O43093 syncerphala
25	744	18.5	1226	1 KFLA_XENLA	P46873 caenorhabdi
26	743	18.4	935	1 K1NH_SYNER	P33176 homo sapien
27	708	17.6	672	1 K1NH_CAEEL	Q61768 mus musculu
28	691.5	17.2	963	1 K1NH_HUMAN	Q12840 homo sapien
29	691	17.1	963	1 K1NH_MOUSE	P21613 loligo peal
30	687.5	17.1	1031	1 K1NH_STRPU	P33175 mus musculu
31	684.5	17.0	1032	1 K1NH_HUMAN	
32	684	17.0	967	1 K1NH_LOLPE	
33	676.5	16.8	1027	1 K1NH_MOUSE	

34	673	16.7	1057	1 EG5_HUMAN	P52732 homo sapien
35	671.5	16.7	815	1 K1NH_CAEEL	P34540 caenorhabdi
36	670.5	16.6	957	1 KFLC_HUMAN	O60282 homo sapien
37	669	16.6	956	1 KFLC_MOUSE	P28738 mus musculu
38	668	16.6	2663	1 CENE_HUMAN	002224 homo sapien
39	651.5	16.2	975	1 K1NH_DROME	P17210 drosophila
40	634	15.7	1056	1 K125_ARATH	P82266 arabidopsis
41	633	15.7	1184	1 B1MC_EMENT	P17120 emericella
42	631.5	15.7	883	1 KLP5_SCHPO	O14343 schizosacch
43	630	15.6	1067	1 EG52_XENLA	Q91783 xenopus lae
44	626.5	15.5	1111	1 KFLP_YEAST	P28742 saccharomyc
45	623	15.5	1060	1 EG51_XENLA	P28025 xenopus lae

## ALIGNMENTS

```

RESULT 1
KFLC_HUMAN STANDARD: PRT; 1103 AA.
AC 043896; 075186;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin-like protein KIFLC.
GN KIFLC OR KIAA0706.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=98352063; PubMed=9685376;
RA Dorrer C., Closek T., Mueller S., Moeller N.P.H., Ullrich A.,
RA Lammer R.;
RT "Characterization of KIFLC, a new kinesin-like protein involved in
RT vesicle transport from the Golgi apparatus to the endoplasmic
RT reticulum."
RT J. Biol. Chem. 273:20267-20275(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176(1998).
CC -!- FUNCTION: MOTOR REQUIRED FOR THE RETROGRADE TRANSPORT OF GOLGI
CC VESICLES TO THE ENDOPLASMIC RETICULUM. HAS A MICROTUBULE PLUS END-
CC DIRECTED MOTILITY.
CC -!- SUBUNIT: MONOMER (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED, WITH MOST
CC ABUNDANT EXPRESSION IN HEART AND SKELETAL MUSCLE.
CC -!- PTM: PHOSPHORYLATED ON TYROSINE.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as the content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U91329; AAC52117.1; -.
CC EMBL: AB014606; BAA31681.1; -.
CC HSSP: P17119; 3KAR.
CC MIM: 603060; -.

```

DR InterPro: IPR000253; FHA\_domain.  
DR InterPro: IPR001752; kinesin.  
DR Pfam: PF00498; FHA; 1.  
DR Pfam: PF00225; kinesin; 1.  
DR PRINTS: PR00380; KINESINHEAVY.  
DR SMART: SM00240; FHA; 1.  
DR SMART: SM00129; KISC; 1.  
DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
DR PROSITE: PS50006; FHA\_DOMAIN; FALSE NEG.  
DR Motor protein; Microtubules; ATP-binding; coiled coil;  
KM Phosphorylation.  
FT DOMAIN 1 355 KINESIN-MOTOR (BY SIMILARITY).  
FT DOMAIN 359 388 COILED COIL (POTENTIAL).  
FT DOMAIN 438 479 COILED COIL (POTENTIAL).  
FT DOMAIN 523 590 FHA.  
FT DOMAIN 633 674 COILED COIL (POTENTIAL).  
FT DOMAIN 828 872 COILED COIL (POTENTIAL).  
FT NP\_BIND 97 104 ATP (POTENTIAL).  
FT CONFLICT 669 669 Q -> R (IN REF. 2).  
FT CONFLICT 955 962 LMGPGRGV -> SGGRGGL (IN REF. 2).  
FT CONFLICT 976 977 NV -> KL (IN REF. 2).  
SQ SEQUENCE 1103 AA; 123071 MM; F14BC398D4B2EC10 CRC64;

Query Match 41.4%; Score 1667; DB 1; Length 1103;  
Best Local Similarity 47.5%; Pred. No. 1.7e-81;  
Matches 354; Conservative 122; Mismatches 157; Indels 114; Gaps 16;

QY 4 GGNKVVVRRPRNARIDGACQIVMEGNOTILPPPGAEKARSGKTMDGPAFA 63  
DB 3 GASVKAVERVRRPFNARETSODAKCVSMOGTWSIINP-----KOSNAPPSFT 51  
QY 64 FDSYVSFEDKNA-PNVRROEDLFODLVPLLDNAFKGYNCCIFAYGOTGSKSYSMAGY 122  
DB 52 FDSYVSHTSTEDPQFASQOQVYVDICEEMILAFEGYNCCIFAYGOTGAKSTTMGRQ 111  
QY 123 K-EHGVIPRICODMFRINELQDKNLTCTVEVSYLEINERVDLLNSTKGNLKYRE 180  
DB 112 EPQGQGIPOCEDLFGRVSEBNQ-SAQLSYSEVSEYMEICERVRLDLPKRSRLRYRE 170  
QY 181 HPSGPRVEDLAKIVASFOEIEMLMDGKARVATNNMETSRSRHAYFTLLTQKWH 240  
DB 171 HPILGPVODLSKLAVSYVDIALDMCGNKARVATNNMETSRSRHAYFTVFTORCH 230  
QY 241 DEETKMDTEYAKISLVLDGSEERATSGATGARKLKEGAEINRSLSLGRVIALADMSS 300  
DB 231 DQLTGDLSEKYSKISLVLDGSEERADSGARGMRLKEGANINRSLTLTGKVISLADMOS 290  
QY 301 GKQKKNOLVPRDSVLTWLKDSIGNSMTAMTAISPADINEETLSTLYRADSAKRIK 360  
DB 291 -KKRKSDFIPRDSVLTWLKDSIGNSMTAMTAISPADINEETLSTLYRADSAKRIK 349  
QY 361 NHAVNDPNAKRMRETKELAOLRSKLQSSGGGGGAGC----- 400  
DB 350 CNALINEDPNARLRELQEEVARKRELMLQGLSASALEGKTEGVSRYGALPAVSSPPA 409  
QY 401 -----SGGPEYESYPPDTPLEKOIVSIQPDATVKKSKAEIYEOLNQSKEIXRD 451  
DB 410 PVSPSSPTTNGELFSPSPPT--ESQI-----GPEAMERLQETEKIITAE 454  
QY 452 NQWEEKLAKTEETIKEREALBELGISIEK--GVGPRYSKEMPHLVNLSDDPLLAEC 509  
DB 455 NEWEEKLAKTEETIKEREALBELGISIEK--GVGPRYSKEMPHLVNLSDDPLLAEC 514  
QY 510 VYNNKPGQTRGVNNOGQAIRLNGSKILKEHCTFEV-----DNVYTYPNKAAMV 564  
DB 515 LYHKKDVTYRGQVDM-----IKLTGQPIRQCHLFRIPQPDGEVYVTLPECGAETV 570  
QY 565 NGVRIIDPTRLRSYRIIGDFHIFERNHPEARAERQGLLRHSTVNSQLGSPAPGRH 624  
DB 571 NGKLVTERLVYAKSGNRIVMGKNHVFRRNHPQALRLER-----GVPPP--- 614

QY 625 DRTLSKAGSDADGDSRSDSPLEPRGKSDMFYARRRASAALIGLDQKISHTYDELAL 684  
DB 615 -----PGPSEPVDMNFAQKLELBO-QGIDIKLE--VEKRLQDL 650  
QY 685 FDDVQKARAVRGLVEPNE---DSDS 707  
DB 651 ENOYRKEKEBADDLLLEOOLYADSDS 676  
RESULT 2  
KFLA\_HUMAN  
ID KFLA\_HUMAN STANDARD; PRT; 1690 AA.  
AC Q12756;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).  
GN KIF1A OR ATRV.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA MEDLINE=96299637; PubMed=8661001;  
RX Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;  
RT "Characterization of a kinesin-related gene ATRV, within the tuberos  
RT sclerosis locus (TSC1) candidate region on chromosome 9q34.";  
RL Genomics 33:421-429(1996).  
CC -!- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC  
CC VESICLE PRECURSORS (BY SIMILARITY).  
CC -!- SIMILARITY: MONOMER (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104  
CC SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X90840; CAA62346.1; -  
DR HSSP: P17119; 3KAR.  
DR MTM; 601255; -  
DR InterPro: IPR000253; FHA\_domain.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR001752; kinesin.  
DR Pfam: PF00498; FHA; 1.  
DR Pfam: PF00225; kinesin; 1.  
DR Pfam: PF00169; PH; 1.  
DR SMART: SM00380; KINESINHEAVY.  
DR SMART: SM00240; FHA; 1.  
DR SMART: SM00129; KISC; 1.  
DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
DR Motor protein; Microtubules; ATP-binding; coiled coil.  
KM KINESIN-MOTOR.  
FT DOMAIN 1 361  
FT DOMAIN 366 383 COILED COIL (POTENTIAL).  
FT DOMAIN 429 462 COILED COIL (POTENTIAL).  
FT DOMAIN 516 572 FHA.  
FT DOMAIN 622 681 COILED COIL (POTENTIAL).  
FT DOMAIN 801 822 COILED COIL (POTENTIAL).  
FT DOMAIN 1575 1673 PH.  
FT NP\_BIND 97 104 ATP (POTENTIAL).  
SQ SEQUENCE 1690 AA; 191083 MM; DBDDEC784624FB4D CRC64;

Query Match 41.3%; Score 1663.5; DB 1; Length 1690;  
 Best Local Similarity 46.7%; Pred. No. 4.7e-81;  
 Matches 351; Conservative 125; Mismatches 167; Indels 109; Gaps 15;

```

OY 4 GGNIVVVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIDGPKARA 63
DB 3 GASVAVAVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIDGPKARA 63
OY 64 PDRSYWSPDKNAP---NARQEDLFDQGLVPLLDNAFKYNNCIRAYCGTSGKSYSMAG 120
DB 52 FDSYWS--HNSPEDINASQKQYRVDIGEEMLQHFEGYNVICIRAYCGTSGKSYSMAG 109
OY 121 YGR--EHGVIPRICODMERINELQKDKNLCTVEVSEIYENRERDLDNSTGKNTLV 178
DB 110 KOEKQGGIIPQLGEDLSRIDTND--NMSYSVEVSEIYENRERDLDNSTGKNTLV 168
OY 179 REHPTGPVEDLAKLVRSFOEINLMDGKNAKARTVAATNNETSSRSNAVFTLTQK 238
DB 169 REHPTGPVEDLAKLVRSFOEINLMDGKNAKARTVAATNNETSSRSNAVFTLTQK 228
OY 229 WHDEETKMDTEVKATISLVDLGSEKATGATGAKTRLEKAGANINSLTGLVYSKLAEM 298
DB 229 RHDAETNTTEKVSRTISLVDLGSEKATGATGAKTRLEKAGANINSLTGLVYSKLAEM 288
OY 299 SSG-----KQKNQLVPRKDSVYTLWLDKSLGNSMTAMIAISPADINEETLSTLRYA 353
DB 289 DSGPKNNKKKKKTDFIPRDSVYTLWLDKSLGNSMTAMIAISPADINEETLSTLRYA 348
OY 354 DSARKIKHNAVNDPPNARMRELKEELAQLRSKLGSSGGGGGAGSGGVEESYPPDT 413
DB 349 DRAKQIRCAVAVNEPNNKRLRELKDEVTLRLDLLYAQGLG-----DIT 392
OY 414 PLEKQIVSIQGDATVYKMS-----KALVEQUNOSEKLYRDLNQW 455
DB 393 DDTNALVGA--SSSSLSLSSRAVSSILHERILFAPGSEEEIERKEKEKTIKAEINTEW 451
OY 456 EEKLAETIEHKEREALEELGISIEK--GFVGPYSHKEMPHLVNDDPLAECLVYNI 513
DB 452 EEKLAETIEHKEREALEELGISIEK--GFVGPYSHKEMPHLVNDDPLAECLVYNI 511
OY 514 KRGQTRGVNODTOAEIRLNSKILKEHCTEN-----VDNVVTIVPEKAAVYNGVR 568
DB 512 KRGQTRGVNODTOAEIRLNSKILKEHCTEN-----VDNVVTIVPEKAAVYNGVR 571
OY 569 IDKPTRLNGVRIITGDPIIFRPNPEEARAROEOSILRHVTSNQLGSPAGRHDRL 628
DB 572 VTEPILNSGNRIIMKSHVETFEPEAKQERER----- 606
OY 629 SKAGSDADODSRSDPLPFRGKDSDFEYARREASAILGLDKISHLTDDLELFDV 688
DB 607 -----TPCARTPAEPYDMAFAQRELEK--QGIOMK--QEMEQRLQLEDEY 649
OY 689 OKARAVRGLVDNEDSDSOSFPVPRDKYMSN 720
DB 650 RRREAEATYLE--QORLDESKLEALQKQMS 680

```

RESULT 3  
 KFLA\_MOUSE STANDARD: PRT: 1695 AA.  
 AC P33173; 061770;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Kinesin-like protein KIF1A.  
 GN KIF1A OR KIF1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95292344; PubMed=7539720;  
 RA Okada Y., Yamazaki H., Sekine-Aizawa Y., Hirokawa N.;  
 RT "The neuron-specific kinesin superfamily protein KIF1A is a unique  
 RT monomeric motor for anterograde axonal transport of synaptic vesicle  
 RT precursors.";  
 RL Cell 81:769-780(1995).  
 RN [2]  
 RP PRELIMINARY SEQUENCE OF 100-247 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93077686; PubMed=1447303;  
 RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,  
 RA Hirokawa N.;  
 RT "Kinesin family in murine central nervous system.";  
 RL J. Cell Biol. 119:1287-1296(1992).  
 CC -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC  
 CC VESICLE PRECURSORS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN  
 CC TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE  
 CC TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE  
 CC CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE  
 CC PERINUCLEAR AND SYNAPTIC REGIONS.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: COMPAINS 1 FHA DOMAIN.  
 CC -1- SIMILARITY: COMPAINS 1 PH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D29951; BAA06221.1; -.  
 DR PIR: E44259; E44259.  
 DR HSSP: P17119; 3KAR.  
 DR MGD: MGI:108391; Kif1a.  
 DR InterPro: IPR000253; FHA\_domain.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR001752; kinesin.  
 DR Pfam: PF00498; FHA; 1.  
 DR Pfam: PF00225; kinesin; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR PRINTS: PR00380; KINESINHEAVY.  
 DR SMART: SM00240; FHA; 1.  
 DR SMART: SM00129; KISC; 1.  
 DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR PROSITE: PS50006; FHA\_DOMAIN; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 KW Motor protein; Microtubules; ATP-binding; coiled coil.  
 FT DOMAIN 1 361 KINESIN-MOTOR.  
 FT DOMAIN 366 383 COILED COIL (POTENTIAL).  
 FT DOMAIN 429 462 COILED COIL (POTENTIAL).  
 FT DOMAIN 516 572 FHA.  
 FT DOMAIN 622 661 COILED COIL (POTENTIAL).  
 FT DOMAIN 801 822 COILED COIL (POTENTIAL).  
 FT DOMAIN 1580 1678 PH.  
 FT NP\_BIND 97 104 ATP (BY SIMILARITY).  
 FT SEQUENCE 1695 AA; 191724 MW; D6ECB88CB8C6C6 CRC64;

Query Match 41.3%; Score 1663.5; DB 1; Length 1695;  
 Best Local Similarity 46.5%; Pred. No. 4.8e-81;  
 Matches 350; Conservative 126; Mismatches 167; Indels 109; Gaps 15;

```

OY 4 GGNIVVVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIDGPKARA 63
DB 3 GASVAVAVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIDGPKARA 63

```

Db	3	GASVAVAVRVPFNSREKRSRKSCKIIMSGSTTTIIVNPKPKET-----PKSF5	51
Qy	64	EDRSYWSVDKAP---NYARDEDFOLGLPBLDNMFKGNINCFAYGQGSKRKSYNMK	120
Db	52	FDYSYWS--HTSPEDINVASQKQYRLRIGDEEMLOHAEIGVNCVIFAYGQGSKRKSYTMKG	109
Qy	121	YK--EHGVIPRICODMFRRINELQOKNLTCTVEVSELYETLYEBRVADLLNPSTKCNLKY	178
Db	110	KQEKDQGGIIIFQLEDLFSRLINDTND--NMSYSVEYSMTIYCEBRDILNPKNKGRLKY	168
Qy	179	REHPSGTGYVEDLAKLVYRSFOEILENIMDEGNKARVYAATNMNETSRSRSHAVETTLTQK	238
Db	169	REHPLIGPYVEDLSKLAVTSYNDIQDLMDSGNKPRVYAATNMNETSRSRSHAVENIIFTQK	228
Qy	229	WHDETKMDTKKVAKISLYDLAGEBRTSTGATGARKKEAEINRSLSITGRVIALADM	298
Db	229	RHDETNITTEKVKISLYDLAGEBRTSTGAKTRELKEBANINRSLSITGKVIASLAEK	288
Qy	299	SSG-----KQKKNOLVPRDPSVLFWLWLDKDSIGSGNSPMAMTAISPADINPEETLSLTRYA	353
Db	289	DSGPKKKKKKKKTDFIPRDSVLTWLLRENLGNSKRYAMTAISPADINDETLSLTRYA	348
Qy	354	DSARRIKNHAVNEDPNPAMRIRELKEBLAOLRSKIQSGGGGGAGSGGSGGVPEESYPPT	413
Db	349	DRAKQIRCNATINEDPNPKLIRELKEDEVTRLDLVLAQGLG-----DLT	392
Qy	414	PLEQIYISIQPDPATVYKMS-----KAEIYBOLNQSCKLYRLDNQTY	455
Db	393	DMTNALVLM--SPSSSLSSLSRAASVSSLHERLIFAPGSSEALIERKETKRIIAELNETW	451
Qy	456	EELKTAETEEIKREBEALAEELGISTEK--GPVGYSHKEMPHIVNLSDDPLAECLVYNI	513
Db	452	EELKLRTPAIMEBEALAEVGMAREDDGLTVGFSPKRTPHLVNINEDPLAECLLYTI	511
Qy	514	KPGQTRGVNQNODTQAEIRLNGSKILKEHCTFEN---VDNVYTVPNKEAAVMYNGVR	568
Db	512	KDGVTRVGRREDAEERODIVLSGHPIKEBHCIFERDSRGGBAVVTLPECGADPTVYNGKK	571
Qy	569	IDKPRRLASGRILLGDPHITRENMHPEBARAEKROESGLNRHSTVNSQLGSPAPGRHRTL	628
Db	572	VTEPSILISNGRIILMGSKSHVFRNHPPOARQER-----	606
Qy	629	SKAGSDADGDSRSDSPLPFGKSDMPVFAAREAAALIGLDQKISHLTDDLEDALEDDV	688
Db	607	-----TPCAETPFAEYVDNAFAQORELLEK--QGDIMK--QEMFORQLEDEDQY	649
Qy	689	QKARAVRGRGLVEDNEDSDQSSFPVVRKYSYN	720
Db	650	RREAREATYLLF--QQRUDYESKLEALQKQDMS	680
RESULT	4		
REFL_MOUSE		STANDARD; PRU; 1816 AA.	
ID	KFIB_MOUSE		
DC	O60575; O9WV65; O9ROB4; O9Z119;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Klimesin-like protein KIF1b.		
GN	KIF1b.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 3).		
RC	TISSUE=Brain;		
KX	MEDLINE=95094296; PubMed=7528108;		
RA	Nagaku M., Sato-Yoshitake R., Okada Y., Noda Y., Takemura R.,		
RA	Yamazaki H., Hirokawa N.;		
RT	"KIF1b, a novel microtubule plus end-directed monomeric motor protein		
RL	Cell 79:1209-1220(1994)."		

FT	DOMAIN	668	737	COLLED COIL (POTENTIAL).	
FT	DOMAIN	841	869	COLLED COIL (POTENTIAL).	
FT	DOMAIN	1702	1799	PH.	
FT	NP_BIND	97	104	ATP (POTENTIAL).	
FT	VASPLIC	289	294	MISSING (IN ISOFORM 2 AND ISOFORM 3).	
FT	VASPLIC	394	434	IDPLIDYSGSGGYUADDFQNNKHRYLVLASRNORGNSTA	
FT	VASPLIC	707	1196	-> T (IN ISOFORM 2 AND ISOFORM 3).	
FT	VASPLIC			YESKLOAIOROVERSLAAETEEEBEEVEEVPWQHEPPLA	
FT				OMARRKKSQHOFTSLRDLDMGNAYIKRANAVSLKELVLO	
FT				FOFVLUPLDILXSPYPPPELLPSMEKTHEDRRPPRYVAVEV	
FT				ODLKNAGATHVSLDKLRLDLMKREMTIRAGEVASSADDS	
FT				ETMTWTGSDPFYDRHWEKLVSSPFIHGLCVBERLADRPSP	
FT				TFSTYADSDITELADBOODAMEDPDEADVDGSPAGTEEG	
FT				SELFSDGHDPEYDRSPMFIIVGRAFYVLSNLIYLPVLIHRY	
FT				AIYSEKGEVREGELVAAYOAIADAEAPAYVSGGSIROSGTAKI	
FT				SDNVEYFQSDPSSAAMTRGSLSEELIYVSGGSSSVYIS	
FT				PPVEYRNANDLDLGSITLDLCKVMAGEESEEIENLKLKLSA	
FT				FEFRVYVLAQAGLILEYADIDICQFNFELRHDEAFSTELPN	
FT				NGRSPLEFYHVONIAVEETEPFVDYIKTKDIVEFVEFH	
FT				-> ADSDGDSDKRSCEESWKLITLSLEKLPKSLQTIYK	
FT				KGLSPSSKREKREPIKQIIPQRRLSKSKWTTIDILKIO	
FT				VKEICYEALNDPFRSROETEKALIVKKELCAAGKGRKDPN	
FT				ERDSMBAYARDVMDTYGDEYDEKIDMTATYGGCGDVYDLKY	
FT				HIDLEDLIDQEVKKONNKDEIKYLRKMLKMEVYVLIG	
FT				SOEOKSOGSHKTEPLVAGANSVDNGSKSESGELGEER	
FT				VSOILMGDPAFRKRLRMWROQIIFKMLQOGLTELKIRO	
FT				NPVPRFTPEENRKPREPFSNPKHNSPSGHHIITDEY	
FT				IELRPIKDEEKRKEKEESOKGVGAASRDYOSAMGTSSD	
FT				ELIWSKOHISNOOPPOLRMRSNLSNNGOQPTTRGTOATYS	
FT				ELIWSHSGHPTADLOTROAKRHINHGRPYCNYNGVOGGS	
FT				STASCCQKQIDRPSHCNOFVTPPRMRKCFSAPIKAGRETT	
FT				MISSING (IN ISOFORM 3).	
FT	VASPLIC	1197	1816	MISSING (IN ISOFORM 3).	
FT	CONFLICT	117	117	G -> V (IN REF. 1 AND 2).	
FT	CONFLICT	520	523	GGTL -> RGDI (IN REF. 1).	
FT	CONFLICT	909	909	P -> T (IN REF. 4).	
FT	CONFLICT	1608	1609	KL -> YW (IN REF. 3).	
FT	CONFLICT	1612	1612	I -> V (IN REF. 3).	
FT	CONFLICT	1784	1784	D -> G (IN REF. 4).	
SO	SEQUENCE	1816 AA;	204080 MM;	E316EC295138E5DE CRC64;	

QY	401	-----SGGPVEESYPDPPLKEQ-----IVSIQDPDAIVK	431
Db	411	DFQNNKHRYLLASENQNRGNSTSMKSLTSS--PSSCSLNSQVGLTSTISIQ--ERIMST	467
QY	432	MSKAIEVQQLNQSEKLYRDLNQTQWEKRLAKTEIHKEREALAEGLISIEK--GEVGPYH	489
Db	468	PGGEARLRTKESKEITIAELNETWEEKLRKTEAIRMEREAALLMEGVAIEDGGTGLVFS	527
QY	490	SKEMPHLYNLSDPDLAECLVYNNKPKQTRGVANNQTOAEIRLNSGKILKEHCTFNV-	548
Db	528	PKRPHPLVNLNEDPLMSECLLYYKIDITRGAQDAERODIVISGAHKEEHCLFRSER	587
QY	549	-----DNVTIYVFNKAAVNVGVRIIDKPTPLRSGRYIILIGDFHIFRNHPPEARAEQO	604
Db	588	SNTGEVITYLTPEPCERSEIYVNGKVAHPVQLRSNRRLIMGNHVFRRNHPQAAHEK-	646
QY	605	SLLRHSVYNSQLGSPAPGRHRTLSKAGSDADGDSRSDPLPHFRGDSDFEYARREAS	664
Db	647	-----TSAEFPSPBPVDTFQAQRELE	666
QY	665	ALILGDK------ISHLTDELAL-----FDVQKARAVRGL-----VED	701
Db	669	K-QGIDMKQEMEKRLQEMELIYKREKEEADLLLEQRLDYESKLLQALQROVETSLAET	727
QY	702	NEDSDSOSSFP	712
Db	728	TEEEEEEVFP	738
RESULT 5			
KF1B_HUMAN	KF1B_HUMAN	STANDARD;	PRT; 1816 AA.
AC	060333;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Kinesin-like protein KIF1B (KIP).		
GN	KIF1B OR KIA0591.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	11		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RA	Bougueleret L., Dufaire-Gare I., Grel P.;		
RT	"DNA encoding a kinesin-like protein (hkip) comprising biallelic		
RT	markers.";		
RL	Patent number WO0063375, 26-OCT-2000.		
RN	12		
RP	SEQUENCE FROM N.A. (ISOFORM 2).		
RA	Yang H.W., Takita J., Chen Y.Z., Soeda E., Piao H.Y., Hashizume K.,		
RA	Hayashi Y.;		
RT	"Cloning of human kif1b gene that maps at 1p36, which is homozygously		
RT	deleted in neuroblastoma cell line NB1 and mutated in some other		
RT	neuroblastoma cell lines.";		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	13		
RP	SEQUENCE FROM N.A. (ISOFORM 3).		
RA	Park W., Shin H., Lee Y.M., Moon E., Choi W., Kim W.;		
RT	"Identification of the human ortholog of mouse Kif1b, a kinesin		
RT	superfamily motor protein.";		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
RN	14		
RP	SEQUENCE OF 479-1816 FROM N.A.		
RC	TISSUE=Brain;		
RC	MEDLINE=98290545; PubMed=9628581;		
RA	Nadalin T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,		
RA	Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. IX.		
RT	The complete sequences of 100 new cDNA clones from brain which can		
RT	code for large proteins in vitro.";		
RL	DNA Res. 5:31-39(1998).		

[5]  
 RN SEQUENCE OF 1449-1816 FROM N.A.  
 RP Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RN VARIANT CMT2A LEU-98.  
 RX MEDLINE=21885926; PubMed=11389829;  
 RA Zhao C., Takita J., Tanaka Y., Setou M., Nakagawa T., Takeda S.,  
 RA Yang H.W., Terada S., Nakata T., Takei Y., Saito M., Tsuji S.,  
 RA Hayashi Y., Hirokawa N.;  
 RT "Charcot-Marie-Tooth disease type 2A caused by mutation in a  
 microtubule motor KIF1B-beta."  
 RL Cell 105:587-597(2001).  
 CC -1- FUNCTION: MOTOR FOR ANTEROGRADE TRANSPORT OF MITOCHONDRIA. HAS A  
 CC MICROTUBULE PLUS END-DIRECTED MOTILITY.  
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY FOUND ON THE VESICLE- AND  
 CC TUBULE-LIKE STRUCTURES WITHIN THE CYTOPLASM (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DISEASE: Defects in KIF1B are a cause of Charcot-Marie-Tooth  
 CC disease type 2A (CMT2A). CMT2A is an autosomal dominant neuropathy  
 CC characterized by normal or slightly reduced nerve conduction  
 CC velocity and axonal loss with little evidence of demyelination or  
 CC hypertrophic changes in nerve biopsies.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AX036604; CAC16629.1; -;  
 DR EMBL: AF257176; AAK49332.1; -;  
 DR EMBL: AY043362; AAK85155.1; -;  
 DR EMBL: AB011163; BAA25517.1; -;  
 DR EMBL: AK022977; BAB14341.1; -;  
 DR MIM: 605995; -;  
 DR MIM: 118210; -;  
 DR InterPro: IPR000253; FHA\_domain.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR001752; kinesin.  
 DR Pfam: PF00498; FHA.1.  
 DR Pfam: PF00225; kinesin.1.  
 DR Pfam: PF00169; PH.1.  
 DR PRINTS: PR00380; KINESINHEAVY.  
 DR SMART: SM00240; FHA.1.  
 DR SMART: SM00129; KTSC.1.  
 DR SMART: SM00233; PH.1.  
 DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 DR Motor protein: Microtubules: Arp-binding; Coiled coil; Mitochondrion;  
 KW Alternative splicing; Disease mutation; Charcot-Marie-Tooth disease.  
 KM  
 FT DOMAIN 1 361 KINESIN-MOTOR.  
 FT DOMAIN 365 386 COILED COIL (POTENTIAL).  
 FT DOMAIN 470 502 COILED COIL (POTENTIAL).  
 FT DOMAIN 556 612 FHA.  
 FT DOMAIN 668 737 COILED COIL (POTENTIAL).  
 FT DOMAIN 841 869 COILED COIL (POTENTIAL).  
 FT DOMAIN 1702 1799 COILED COIL (POTENTIAL).  
 FT NP\_BIND 97 104 ATP (POTENTIAL).

FT	VARSPPLIC	289	294	MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	VARSPPLIC	394	434	IDPLIDYSGSGSKYTKDFPNKNNKRYLASENORPHFBSRA
FT	VARSPPLIC	707	1196	-> T (IN ISOFORM 2 AND ISOFORM 3). YESKLAQOVETRSIAETTEESEEEEPWTOHEPFLA OMAFRRKSHQFSLDLMLGNVAFLKEAIAISVELKRVQ FOVLLITLTLSPPELLPTENAKETHEDPPPTVAAYV ODLNGATHWLSLEKRLDRLKREMYDGEVMASSADPS ETTVGSDPEYDRHMFKLGVSSPIFGHCVENLADPTSP TSTADSDITELADROODEMEDPDDPDADGASCTEBS SDFPSDHPDFTRSFHILVGRAFYLSNLILPYVLIHAY AIVSEKGVGFLVAVOALAEADPDVSGIROSQTAL SPDNEYFNOSDSSVAMTRSGLSLEELRIEYEGOSSEVLT PPEEISRLINDLKSLLDGGKMEGFESEIENHKLGA FTRVYVLOASGILPEADIECOFNHRHDEAFSTPELKN NGRSGPLAAYNONIAVEITRESFVDTIKTPYFVEVGH -> ADSDSGDSDSKSCSESWKILTSIREKLPSKLOTYIK KGLPSGGRREPIMYQIPORRRILSKMSKWTISDLKLOA KEICEYVALNDRHSROEIELALIVMKELCAMYGRKDPN ERDSWRAVARDVDTGVGDEKIEDVATKGSSTDVLDLY HIDKLEIDLOEVKKNNMKDEIKVLRNKLMEKYLPLG SOEOKSPGSHAKPEYACAGYSTSENVSXGDNGLAKBER VSOLMNGDPFRGRRLRMHROEDIRFNLDQOELITOLRQ NVPHRFTIPENRPREPEKSNPHRNSWSTGTHILTEDEV IELRIPRDEARKNKEESQEGKGFKDPGPGSGQGR SODHIOYSKOHINNOOPOLRMRSNLNNGOPKSTRCOAS ASAESLNSHSGHPTADYOTFOARKRHITHOROSYCNVTGG LEGNAATSYOKORJDKSHCSQFVTPPRMRQPSAPNLKACR ETTV (IN ISOFORM 3). MISSING (IN ISOFORM 3). Q -> L (IN CMT2A). /FTID=VAR_011515. TNH -> KIN (IN REF. 1). D -> E (IN REF. 1). R -> L (IN REF. 1). VYV -> AVF (IN REF. 1). ILATV -> NLSTE (IN REF. 1). T -> I (IN REF. 1). D -> E (IN REF. 1). I -> N (IN REF. 1). SQ SHQUNCE 1816 AA; 204427 MW; ECAB9647215AC7DB CRC64;

Query Match 40.3%; Score 1623.5; DB 1; Length 1816;  
 Best Local Similarity 44.1%; Pred. No. 7.1e-79;  
 Matches 349; Conservative 127; Mismatches 178; Indels 137; Gaps 16;

QY	4	GGNIKYVVRVRFNAREIDGAKCIYVMENQTLTPPCAEKARSGKTIMGPAFA 63
DB	3	GASYKVAVRVRFNAREIDGAKCIYVMENQTLTPPCAEKARSGKTIMGPAFA 63
QY	64	FDRSYWGF-DKNAPNVAROEEDLPDGLVPLIDNFKYNNCIFYAGGTGSGKSYSMGYG 122
DB	52	FDRSYWGF-DKNAPNVAROEEDLPDGLVPLIDNFKYNNCIFYAGGTGSGKSYSMGYG 122
QY	123	KEH---GVIPRICODMFRINELQKMLTCTVEVSEYIEINERYRDLINSTKGNLKVRE 180
DB	112	EEGQAGIIPOLCELF-ETNNDCNENEMSVSEVSEMEICERADLLNPKKNCNLVRD 170
QY	181	HPSTGPYVEDLAKVNSFOEITENLMDGKARFVAATNNMETSSRHAVFTLTQKWH 240
DB	171	HPSTGPYVEDLAKVNSFOEITENLMDGKARFVAATNNMETSSRHAVFTLTQKWH 240
QY	241	DEETKMDTEYAKISLVNGLSGERASTGTGATLKGAEINRSLTGLGVIALADM-- 298
DB	231	DNETILLATVAKSVKSLVDLGSDDRAASTGAKGTAKGATILNLSLTGLGVIALADM-- 298
QY	299	---SSGROKKNQOLVPRYDSVLTWLLKSLGNSMTAMIAISPADINEETLSTLRVADS 355
DB	291	CTSKSKKKKKTDPLPYRDSVLTWLLKSLGNSMTAMIAISPADINEETLSTLRVADS 355
QY	356	AKRIKHAHVNEDPNAMKIRELKEELAOULNSKLOSSGGG-----GGAG----- 399
DB	351	AKRIKHAHVNEDPNAMKIRELKEELAOULNSKLOSSGGG-----GGAG----- 399



QY 400 -----GSGGVESYPPDPLEKQ-----IVSIQDPDATYK 431  
Db 411 DEQNKHRYLLASENQRFHSTASMGSLTSS-PSSCSLSSQVGLTSTSIQ-ERINST 467  
QY 432 MSAKATVQLQNOSELYLDLNOTWEEKLAKTEIHKEREALEELGISTEK-GVGPYPH 489  
Db 468 PGGEBAIRLEKSEKTIYELNETWEEKLRKTEALIMERALLAENGVAIREDGGTIGVFS 527  
QY 490 SKEMPHLVNLSDDPLLAECIYVNIKPGOTRVGNVODTOAEIRLNGSKILKEHCFEYV- 548  
Db 528 PKKPHVLVNMEDPLMSSECLLYITIDGITRVQADAEKRODVLSCAHKEKHCIFRSER 587  
QY 549 ----DNVTIVPNEKAAVNNGVRIDKPTRLSGYRIILGDEHIFRPNPEBARAEQO 604  
Db 588 SNSGVITVTLPEPCERSEETVNGKRVSPQVLRSNGRIIMGNHVRFRPNPEQARAERK- 646  
QY 605 SLLRHSVNSQLGSPAPRHRTLSKASDADGSRSDSPILPHFGKSDSWFYARREAS 664  
Db 647 -----TPSAETPSEPYDWTFAQRELE 668  
QY 665 AILGLDQK-----ISHLTDELDALEFDD-----VOKARAVRGGLVED 701  
Db 669 K-QGIDMNGEMEKRLQEMELLYKKKEKEEADLLLEQQRLDYESKLDALQKQVTRSLAET 727  
QY 702 NEDSDSQSSFP 712  
Db 728 TEEEEEVEVP 738

RESULT 6  
KPID\_RAT STANDARD; PRT: 1097 AA.  
ID KPID\_RAT STANDARD; PRT: 1097 AA.  
AC 035787;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE kinesin-like protein KIF1D.  
GN KIF1D.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid-10116.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Wistar;  
RA Rogers K.R., Griffin M., Brophy P.J.;  
RT "The secretory epithelial cells of the choroid plexus employ a novel  
RT kinesin-related protein.";  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AJ000696; CAA04248.1; -.  
DR HSSP; P17119; 3KAR.  
DR InterPro; IPR000253; FHA\_domain.  
DR InterPro; IPR001752; kinesin.  
DR Pfam; PF00498; FHA.1.  
DR Pfam; PF00225; kinesin.1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00240; FHA.1.  
DR SMART; SM00129; KISC.1.  
DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; FALSE\_NEG.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.

DR PROSITE; PS50067; FHA\_DOMAIN; FALSE\_NEG.  
KM Motor protein; Microtubules; ATP-binding; Coiled coil.  
FT DOMAIN 1 357 KINESIN-MOTOR.  
FT DOMAIN 358 380 COILED COIL (POTENTIAL).  
FT DOMAIN 437 478 COILED COIL (POTENTIAL).  
FT DOMAIN 520 587 FHA.  
FT DOMAIN 630 671 COILED COIL (POTENTIAL).  
FT DOMAIN 824 868 COILED COIL (POTENTIAL).  
FT NP\_BIND 96 103 ATP (POTENTIAL).  
SQ SEQUENCE 1097 AA; 122333 MW; 8FA0B1C7579BA5B CRC64;

Query Match 40.1%; Score 1617; DB 1; Length 1097;  
Best Local Similarity 45.9%; Pred. No. 7.9e-79;  
Matches 357; Conservative 129; Mismatches 184; Indels 108; Gaps 20;

QY 4 GGNIRYVVRPFPNAREIDRGAKCIYRMGNCTIILPPGAEEKARKSKITMDGKARA 63  
Db 3 GASVKAIVRVRPFPNAREISQDAKCVSMGNTSTIINP-----KQSRMPL--KA-S 50  
QY 64 FDRSYWSEDK-NAPYARQEDLFODLGVPULDNAEFGYNNCIFYAGQOTGSKSYMMGYG 122  
Db 51 FDRSYWSHTSVEDPQFASQOQYVRDQGEMLLHAFEGVVCIFYAGQOTGAGSYTMGRQ 110  
QY 123 K--EHGVIPRICQDMFRINELQKDKNLTCYEVSYLEYINERVRDLNLPSTKGLKYRE 180  
Db 111 EPGQGIYVQLCEDLFESRYN-VNQSALSYSEVSMEYICERVRLDLPKRSGLRYRE 169  
QY 181 HPSGTGYVEDLAKLVRSQOEIENLMBEGNKARVAAATMMNETSSSHAVFTLTQKH 240  
Db 170 HPILGPYVODLSKLVATSYADIADLMDCGNKARVAAATMMNETSSSHAVFTIVPQRSH 229  
QY 241 DEFKMDTEKVAKISLVDAJGSEATSTGATGARLKEGAEINRSJSTLGRVTAALADMS 300  
Db 230 DDLTGLDSEKYSKISLVNLAGSERADSSARGMRLEKGINIKSLTTLTKVTSALADLOS 289  
QY 301 GKQKNQVLVPYRDSVLTWLLKDSLGNSMTAMIAISPADINFEETLSFLRYADSAKRIT 360  
Db 290 -KKRKSDFIPYRDSVLTWLLKDKNGNSRTAMIALSPADINVEETLSFLRYADRTKQIR 348  
QY 361 NNAVNEPDNNAWIRLEKLEQLNRKLSQSGGGGAG-----SGG----- 403  
Db 349 CNAVINEPDNARLIRLEQEVARLRELLMAOGLSASALGGLVEEGSPGVLPAASSPPA 408  
QY 404 PYEESYPPDPLEKQIVSIQDPDATYKMSKAEIYQLNOSELYLDLNOTWEEKLAKTE 463  
Db 409 PASPSFP--PINGLEPEPSPSAE-POIGPEAMERLOETEKIYALNETWEEKLAKTE 465  
QY 464 EIHKEREALLELGISIEKGVGYPHSEMPHLVNSDDPLLAECIYVNIKPGOTRVGNV 523  
Db 466 ALRMEREALLAEMGSPGWRIVGVSPPKTPHVLNEDPLMSECLYHIKDGVTREVGV 525  
QY 524 NODTOAEIRLNGSKILKEHCFEYV----DNVTIVPNEKAAVNNGVRIDKPTRLRS 578  
Db 526 DVD---IKLTOQFIREQCLFRSTIPQPDGEVAVVLEPEGAEYVNGLVTEPLVLSG 581  
QY 579 YRIILGDFHIFRPNPEBARAEQOSILRHSVTNSQLSPAGRHDRFLRSAGSDADG 638  
Db 582 NRIYVGNKNVFRPNPEQARLERER-----GVPPP----- 611  
QY 639 SRSDDPLPHFRCKDSDFYARREASAILGLDQKISHLTDELDALEDDVOKARAVRGL 698  
Db 612 -----PGPSEPYPVDMNAQKDWLEQ--OGIDIKLE--NEKRLODENQYRKEKEADLL 661  
QY 699 VE-----DNEDSDSQSS-----FPVBDKYWSNCTINFSDTATITPGPRRS 740  
Db 662 LEOQRLYADSDSGEDSKKSCESWRLLISLKDLPN-----TVQTTVKRCGLPSS 713

RESULT 7  
ID U04\_CABEL STANDARD; PRT: 1584 AA.  
AC P23678;







[illegible]

DR EMBL; Z26077; BAA05070.1; -.  
DR HSPB; P17119; 3KAR.  
DR MGJ; MGI:107688; KIF3B.  
DR InterPro; IPR001752; Kinesin.  
DR Pfam; PF00225; Kinesin; 1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00129; KISC; 1.  
DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Neuropeptide; Kinesin-motor (by similarity).  
FT DOMAIN 1 345  
FT DOMAIN 346 579  
FT NP\_BIND 580 747 GLOBULAR.  
FT NP\_BIND 96 103 ATP (POTENTIAL).  
FT DOMAIN 386 393 POLY-GLY.  
FT DOMAIN 394 405 POLY-GLU.  
FT DOMAIN 723 730 POLY-SER.  
SO SEQUENCE 747 AA; 85288 MW; FA369A4190EC8B47 CRC64;

Query Match	21.08	Score 844.5	DB 1	Length 747
Best Local Similarity	32.98	Pred. NO. 6.6e-38		
Matches 256	Conservative 116	Mismatches 253	Indels 153	Gaps 27

OY	6	NIKVYVRPRPNAE----	IDRAKCIYRMBGNOTILTPPGAEERKRSKGTIMDCPKA	61
Dd	9	SVRVVRCRRPNMGKEKAASYKVDVUVKLT--	GQVSVPNKPG-----TSHEMRPT	56
OY	62	FAPDRSY-MSPDKNAPNVAROEDLFODLGVLLMNAFKGYNNCIFAYGONGSGSVMNG	120	
Dd	57	FTFDVIVDMN-----AKOFELDEFRPLVDSVLQGEFGTIFAAGTGTTGYTMEG	108	
OY	121	Y---GKEHGVIPRICODMFRRINIELOKDKNLTJCHEVSEIYENERYNRDILLNPSTGNKL	177	
Dd	109	VRGDERKGVIAPNSPHDIHTFHSISQNOQL---	VRASILETYQEBLRDDLSKSQTRLE	165
OY	178	VREHPSTPIYVEDIAKLVYBSFOEIENMLDGKARFYVAATNNETSRSRHAVFTLLTQ	237	
Dd	166	LKERDPITGVYVKDISSFYTKSVKTEIEHMANNONRSYGATNNMHSSSHAIPIVTI--	222	
OY	238	KWHDETBMQDE---KVAKISLVDLAGEBRTSGATGARLCBEAAELNRSLSITGRVIAA	294	
Dd	224	--ECSEVGLDGENHIRGKGLVNDLAGSERAKTAGOERKEATKTNLSIALGANVISA	281	
OY	295	LADMSGROKXNOLVAPYDSVLTMLLKSLLOGSNMTAMIAISPADIINFETESTLYRAD	354	
Dd	282	LVD-----GSTHIPTYDSKTLTFLDOSLGMAKTVMANVGASVNEBTTLTTLKAN	335	
OY	335	SAKRITHNAVUNEDPNARMIRELKEBELAQLRSKIOSGGC-----GGAGSGSGP	404	
Dd	336	RAKNINKPNRYNEBDPKALLREFPEBIIRALKOLEKRSIGRRKRREERGCGGGCEE	395	
OY	405	VE---ESRPPOTPLEKOVSLOOPDATVKKMKSAEIVE-QLNQSEKLYRLNOTWMEKL	455	
Dd	396	EEEBEGEEDGDDDDYWRBOEKLELEKRA---IVEDHSLVAEEKM---RLKEKE	447	
OY	460	AKEBIEHKEREBAALLETIASIE---KGFVPHY---HSEKMPHLVNLSDPLLAECLVY	511	
Dd	448	KKMDELRRERKDA-EMGAKIKAMEBKSKLVGGKNIYDHITNQOKILLOKROELAO-----	502	
OY	512	NIKPGOTEVGNVNODTOAEIRLNGSKLLEKHCTENDNVNTIVPNEKAAVMNGVIDK	571	
Dd	503	----KRREREIOQ--QMESDEETLEKE--TVTSLQOEVDI-----KTKK	540	
OY	572	PTRLASGRILLGDPHIRFPNHPEARAAREBOS-----LLKRSVINSOLG-	617	
Dd	541	LKLEFSKLOAVKAEIHDOEBIEROLETOETONELTRELIKHLIIENFIPEEKRMIM	600	
OY	618	-----SPARGDRRTLKSGSOADPSRDS-----PLPFGRKDS	655	
Dd	601	NRSFPDDEEDHHKLHPTRLLENQQMMKRPVAVGKYRPLSHAAKSMKIRPEPYRAEN-	655	
OY	654	DMFYAREBASAILGLDCKISHTLTDDELDALEFDVOKARAVRGLVEDNEBDSQSASF	711	

Db 660 -----IMLEEDMPRTTDEYEGPAISPKVQ--AALDAALQDEDEIDYDASSF 705

RESULT 12

K121\_STRPU STANDARD; PRT; 742 AA.

ID K121\_STRPU STANDARD; PRT; 742 AA.

AC P46871;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).

GN KRP95.

OS Strongylocentrotus purpuratus (Purple sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;

OC Strongylocentrotus.

OX NCBI\_TaxId=7668;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Egg;

RX MEDLINE=94050179; PubMed=8232586;

RA Cole D.G., Chinn S.W., Medman K.P., Hall K., Vuong T.,

RA Scholey J.M.;

RT "Novel heterotrimeric kinesin-related protein purified from sea

RT urchin eggs.";

RL Nature 366:268-270(1993).

CC -1 SUBUNIT: PRIMER OF A 115 kDa SUBUNIT AND TWO KINESIN-LIKE SUBUNITS

CC OF 95 AND 85 KDA.

CC -1 PTM: THE N-TERMINUS IS BLOCKED.

CC -1 SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN

CC II SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U00996; AAA87393.1; -.

DR HSSP: P17119; 3KAR.

DR InterPro: IPR001752; kinesin.

DR Pfam: PF00225; kinesin; 1.

DR PRINTS: PR00380; KINESINHEAVY.

DR SMART: SM00129; KISC: 1.

DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil.

FT DOMAIN 1 337 KINESIN\_MOTOR (BY SIMILARITY).

FT DOMAIN 338 613 COILED COIL (BY SIMILARITY).

FT DOMAIN 614 742 GLOBULAR (BY SIMILARITY).

FT NP\_BIND 95 102 ATP (POTENTIAL).

SQ SEQUENCE 742 AA; 84202 MW; 47C40A367BA77B5 CRC64;

Query Match 20.5%; Score 827.5; DB 1; Length 742;

Best local similarity 31.5%; Pred. No. 5, 2e-37;

Matches 250; Conservative 122; Mismatches 286; Indels 135; Gaps 25;

QY 181 HPSTGYVEDLAKLVYRSFQEIENLMDENKARTVAATNNNETSSRSHAVFTLLTQKWH 240

DB 168 RPDGYVVKDLSFVTKRSYVEIHMVTGVNNNSVGSNTNMNHSRSHAFITTI-----E 223

QY 241 DEETKMDTE---KVAKISLVDLGSEKATSTGATGARKGAGATNSLSLGVATIALAD 297

DB 224 CSELGVADGENHIEVGRKLNLDLAGSEKQATGATGTRKLENTKINISLSLGVATIALAD 283

QY 298 MSSGKQKKNOLVPRDSVLWKLKSLGNSMTAMIAISPADINFEETSTRTYADSAR 357

DB 284 -----GKSHITPRDSKLTLLRLQDSLGNAKTYVMANMPASYNDETITTLRYANRAK 337

QY 358 RIKNHAVNEDPNAMIRLEKEELAQSLRSKLOSSG--GGGGAGSGGPGVEEYPPPTPLE 416

DB 338 NIKNKPINEDPDALREFQVEISRLKQALDKKPSDGRKKKKRPGGGDDIDEDE 397

QY 417 KOIVSTQOPDPAIVYKMSKATVEQLNOSKLYLDLNOTWEKTLAKTEI-----H 466

DB 398 TEEEGDEMEDEEYKESQKLEE--EKEMINQSKIAEKKOKLSEYOKRQGEIKKEH 454

QY 467 KERPAALKEELGISLEKGFVGPY-----HSKEMPHLVNLSDDPLAECLVYNIKPGQTRVGN 522

DB 455 QQKEMLEKRTKAKESKLVGKSTVDHTNEQOKRIE--EQRLLLAE-----EKN 501

QY 523 VNDDQAEIRLNGSKILKECHTEFENDVNTIVPNEKAAYVNGVRIDKPTRLRSGRYRI 582

DB 502 RERDMERKLEQDDKQVEIEIGTFESSLQOEVE-----VKTKKLKLFKALQSY 548

QY 583 LGDFHIFRHNHPEAR--AEQEOSILRHSTNSOLSGPARGHRT--LSKASDADGDS 639

DB 549 KSDIQDLQDDHAREROLEOTQNELRELKRVADNFIPEVERKTTTTRAVFDEETEE 608

QY 640 RSDPLPFRGKD-----SDWFYAR-----RPAASAIILGLDOKIS 674

DB 609 WLTPPLAKAGPSQMAKRPVSAYGNRRPAD--YARAADMGCPKRAKNILISVLDMP 666

QY 675 HLT--DDELDAFLDDYOKARVRGLVEDNEDSDSQSFPV-----BDKYMSNGTID 724

DB 667 NRTTRDYEGSVAPRVQ--AALDAALQDEDDDLLEVQPEVFAKTKLKDKVRSK---- 719

QY 725 NFSLDATITMPTG 737

DB 720 ---HKAVAKPGS 728

RESULT 13

FL10\_CHLRE STANDARD; PRT; 786 AA.

ID FL10\_CHLRE STANDARD; PRT; 786 AA.

AC P46869;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Kinesin-like protein FLA10 (Khp1 protein).

GN FLA10.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI\_TaxId=3055;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=137;

RX MEDLINE=94299638; PubMed=8027176;

RA Walther Z., Vashishtha M., Hall J.L.;

RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous

RT protein.";

RL J. Cell Biol. 126:175-188(1994).

CC -1 FUNCTION: PROBABLY INVOLVED IN FLAGELLAR ASSEMBLY AND MAINTENANCE.

CC MAY PLAY A ROLE IN FLAGELLAR SYNTHESIS.

CC -1 TISSUE SPECIFICITY: FLAGELLAR AXONEME.

CC -1 SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN

CC II SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: L33697; AAA21738.1; -  
DR HSSP: P17119; 3KAR.  
DR InterPro: IPR001752; Kinesin.  
DR Pfam: PF00225; kinesin; 1.  
DR PRINTS: PR00380; KINESINHEAVY.  
DR SMART: SM00129; Kisc. 1.  
DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE: PS0067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; Microtubules; ATP-binding; coiled coil.  
FT DOMAIN 1 358 KINESIN\_MOTOR (BY SIMILARITY).  
FT DOMAIN 367 687 COILED COIL (POTENTIAL).  
FT DOMAIN 688 786 GLOBULAR (POTENTIAL).  
FT NP\_BIND 97 104 ATP (POTENTIAL).  
FT DOMAIN 388 391 POLY-GLY.  
FT DOMAIN 705 714 POLY-GLY.  
FT DOMAIN 756 759 POLY-ASP.  
SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 20.5%; Score 825.5; DB 1; Length 786;  
Best Local Similarity 39.3%; Pred. No. 7.2e-37;  
Matches 204; Conservative 80; Mismatches 168; Indels 67; Gaps 14;

OY 2 SGGG--NIKVVVRVPPNAREIDRGAKCYRME---GNOTITPPPGAEERKARKSGKTM 56  
DB 4 AGGGSESVKVVVRCPLNGKEKADRSRLVDMVDAGOVKNPPADASEP----- 54  
OY 57 DPKAFAPDRSY--WSPDKNAPYARQEDLPDLGVPLDPAKFGYNNCFAYGOTGSGKS 115  
DB 55 --KAFITPDQYVDWV-----COQDVFDITARPPLIDSCLEGNGTIFAIGOTGTGMS 104  
OY 116 YSMGYSYGE--HGVIPIRICODMFRINELQDKMLCTVEVSYLEYNERVRDLNPST 172  
DB 105 HTMEGKDEPPELRGLIPMFERYVFELIARDSGTKEF--LVRSSELEYNEVRDLNGKH 162  
OY 173 KGNLVREHPSTGPRVVEDLAKLVRSFOEIENTLMDGKNARKYVAATNNMETSRSRAVVT 232  
DB 163 SKKMKELSPDQGVYVKDLSQFCVKNYEEMNRYVLGAKDNROYGATLNMQDSRSRHSIPT 222  
OY 233 LTL-----TQMHDEETKMDTEKAKISLVLDLAGESEATSTYGATGARLKEGAEIN 282  
DB 223 ITIECTIKESAAAKPQAKKDSNHYRGKLVLDLAGESEKQDTGATGDKLKEBKITIN 282  
OY 283 RSLSTLGRVIALADMSGCKOKNQLVPYRDSVLTWMLKSLGNSMTAMIAISPADIN 342  
DB 283 LSLTALGNVIALYDQKSGH-----IPYRDSKLTFRLLQDLSLGGMTKYVMANITGPADMN 336  
OY 343 FEETISTLRYSARIRKIHNAVNNEDPNARMTRELKEELAQRSKIQSGGGGGG---A 398  
DB 337 YDETSTLTARAKRANIOKPKINEDPKDAMLRQFOEETIKLKEQLAANAAGSGGHTMP 396  
OY 399 GSGSGPVE-----ESYPPD-----TPLEKOIVSIQODPAVYKKSKAEIYEQUNO 443  
DB 397 SGGSGPPTOKIYERTEVDPPDIDAIKAQMAELDAKKKSISTEALDKAREEAANAAR--KQ 455  
OY 444 SEKLVRDLNQWTEERLAKTEETIKEREALAEELGISIEK 482  
DB 456 LQAIIDQCKTEAOKKARDAALKQAEEARALAG--AIKK 493

RESULT 14  
KINL\_LEICH STANDARD; PRT; 955 AA.  
AC P46865;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Kinesin-like protein K39 (Fragment).  
GN KIN.  
OS Leishmania chagasi.  
OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5669;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MHOM/BR/82 / ISOLATE BA-2;  
RX MEDLINE=93133867; PubMed=8421715;  
RA Burns J.M. Jr., Shreffler W.G., Benson D.R., Chailb H.W., Badaro R.,  
RA Reed S.G.;  
RT "Molecular characterization of a kinesin-related antigen of  
RT Leishmania chagasi that detects specific antibody in African and  
RT American visceral leishmaniasis."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).  
CC -1- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.  
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: L07879; AAA29254.1; -  
DR HSSP: P17119; 3KAR.  
DR InterPro: IPR001752; kinesin.  
DR Pfam: PF00225; kinesin; 1.  
DR SMART: SM00129; Kisc. 1.  
DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE: PS0067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; Microtubules; ATP-binding; coiled coil; Repeat.  
FT DOMAIN 1 398 KINESIN\_MOTOR (BY SIMILARITY).  
FT DOMAIN 426 >955 COILED COIL (POTENTIAL).  
FT NP\_BIND 122 129 ATP (POTENTIAL).  
FT DOMAIN 704 >955 7 x 39 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 704 742 1.  
FT REPEAT 743 781 2.  
FT REPEAT 782 820 3.  
FT REPEAT 821 859 4.  
FT REPEAT 860 898 5.  
FT REPEAT 899 937 6.  
FT REPEAT 938 >955 7 (PARTIAL).  
FT NON\_TER 955  
SQ SEQUENCE 955 AA; 106168 MW; 8CA76815BE84C6B9 CRC64;

Query Match 20.2%; Score 814; DB 1; Length 955;  
Best Local Similarity 28.1%; Pred. No. 3.9e-36;  
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

OY 7 IYVVVRVPPNARE--IDRGAKCYRMEGNQITLPP-----PGAERKARKSGKTMND 57  
DB 13 YVSVYVRPDLNRENNAPGRTKVTAQAQAAVTVKVLGSGNSGAASMGARRAVD 72  
OY 58 GKAPAFDQSYSP---DKNAPRYARQEDLPDLGVPLDPAKFGYNNCFAYGOTGSGK 114  
DB 73 ----TQFDHVFWSVETPDACATPATQADVFRITGVPLOVHAFFDGFNSCLFAYGOTGSK 128  
OY 115 SYSMNG-----YGKEHGYIPRICODMFRINELQDKMLCTVEVSYLEYNERVRDL 168  
DB 129 YTTMAGADVASLSEGNNGTTPRICLEIFARKASVEAQSGSRMTVELGYEVNERSDLL 188  
OY 169 NPSTKG-----NLKVRHPSTGPRVVEDLAKLVRSFOEIENTLMDGKNARKYVAATNN 220  
DB 189 GRRKGVKSGGEVYVDVREHPSRGVFLGQRLVEGSLDDVVRILTEIGNVYRHTASTKM 248  
OY 221 NETSSRSHAVFTLLT-----TQMHDEETKMDTEKAKISLVLDLAGESEATSTYGATGARL 276  
DB 249 NDRSSRSHAIIMLLKEERTMTTKSGETIRTAGKSSRMNLVLDLAGESEATSTYGATGARL 308

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@sib.slb.ch).

EMBL; D12645; BAA02166.1; -.  
PIR; B44259; B44259.  
DR HSSP; P17119; 3KAR.  
DR MGD; MG1:107689; Klf3a.  
DR InterPro; IPR001752; kinesin.  
DR Pfam; PF00225; kinesin; 1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SMO0129; KISC; 1.  
DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Neutone.  
FT DOMAIN 1 350  
FT DOMAIN 351 586  
FT DOMAIN 587 701  
FT NR\_BIND 100 107  
FT DOMAIN 442 445  
FT POLY-GU.  
FT POLY-ARG.  
SQ SEQUENCE 701 AA; 80167 MW; 2405872DF2D85A29 CRC64;

Query Match 19.9%; Score 800; DR 1; Length 701;  
Best Local Similarity 31.8%; Pred. No.1,4e-35;  
Matches 231; Conservative 110; Mismatches 231; Indels 154; Gaps 25;

DQ 6 NIKVVVRVPPNAREID---RGACIVMEEGNOTILPPPAAEEKKARKSGKTIMGPKAF 62  
|||:|||::||||::|:  
14 NVKVVRRCRLNEREKSMCYROAVSDMERGTIY-----HKTSSN----EPRTTF 61  
|||::|:  
DY 63 AFDDSYSEFDKNAPRYAQEDLFODLGVPLLDNAFEKYNNCIFAYGOTGSKTSMMGYG 122  
|||::|:  
Db 62 TFDTVF-----GPE-SKDLDVTNLTPARDISLYEIGNCTIFAYGGTGCKTFTEMGVR 114  
|||::|:  
DY 123 KE---HGVIPIQCDDMFRIENELOKDKKLCTVEYSLEYIENERVDLRDLNSTGNLKVR 179  
|||::|:  
Db 115 AVPGIRGVIPNSFAHIIEPHIAKEGDIREF--LVKYSTLEITINEEVRDLLKGDOQRLEVK 172  
|||::|:  
DY 180 EHPSSTGPVEDLAFLVRSFOEIENLADDEGNKAKPTVAATNNETSSSRSHAVFTLT--TQ 237  
|||::|:  
Db 173 ERPVGVIYIKPLDSAYVMNADMORIMTLGRKNRSVCGTANNESHSRHAIFTITICSE 232  
|||::|:  
DY 238 KWHDDETAMDIEKAVALSLVLASERATSGANGARLKEAELENRSLSSTIGRIATALAD 297  
|||::|:  
Db 233 KCVGNNGHV---RNGLHLVDLAGSERAKTAGTGORKRETKNTLSLSTLGNVISALVD 289  
|||::|:  
DY 298 MSSGKOKKQLVPYRDVYLWLKDSLLOGNSMTAMIALISPADINFEBETISTLRPADASK 357  
|||::|:  
Db 290 -----GSTHFVPRNSKLTRLTDSDIGSGNKTYMCANICGPADYNDETISTLRYANRAK 343  
|||::|:  
DY 358 RIKNHAVVNEDPNRMRIRELKEBELAOLRSKIQQSGG--GGGAGCGSGPVPE----- 407  
|||::|:  
Db 344 NIKKHARTNEPDKALLNQFKETEELKKLLEGEEVVSQSISGEEDDERGELGEDGEK 403  
|||::|:  
DY 408 ----SYDPDTPLEKO-----IVS 421  
|||::|:  
Db 404 RKRRRDDAGKKKVSPPDKMVEQAIDEERRALKFLDMEEBERRKKAARALERRRKDLLKA 463  
|||::|:  
DY 422 IQCPDAIVVKMSKAE-----IVQLNSEKLYNDLNQTWEKLAITEIHKRBBAA 472  
|||::|:  
Db 464 QOEHQSSLFEKLSALEKKRYIVGVDDLAKAKEOEFKILEESNMLEEERRRRRAEQLEKEEK 523  
|||::|:  
DY 473 -----LEELGISLEKGFVPGYHSKEMPHLVNLSDDPLAECLVYNNLPKPOTRGANNOD 526  
|||::|:  
Db 524 EQEMLDIEKTSIQBERQG--KTKKLKWKTM-----LMAKSSEMDAQOE 568  
|||::|:  
DY 527 TOAIR-LNGSKLIKHCFFEN--VINNV-----TIYPN-----EKAIVMV 564



Db 569 HOREIEGLELLENIRQLSREIRLQMLIIDNFIPODYQEMIENTVHMNEDIGEWJLKCVAITG 628  
OY 565 NGVRIDKPT---RLRSGYRIIGDFHIFRFNHPPEARAROEOSILRHSTNSQLGSPAP 621  
Db 629 NNMKQOTVPDPKKEKRPPEVDLS--HY-LAYTEESL--ROSLMKLERPRITSGKARPERT 683  
OY 622 GRHDRT 627  
Db 684 GRRKRS 689

Search completed: July 25, 2002, 05:46:59  
Job time: 448 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 05:44:26 ; Search time 70.23 Seconds  
(without alignments)  
1931.198 Million cell updates/sec

Title: US-09-235-416-1

Perfect score: 4030  
Sequence: 1 MSGGNIKVVVRVRFNARE.....ELRQOQOMBEALTKAKQEF 784

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_proteint:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1635	40.6	1773	5	09V776
2	1630.5	40.5	1671	5	09V776
3	1590.5	39.5	1581	5	018778
4	1525.5	37.9	2205	5	09NGQ2
5	1468	36.4	1749	11	09ECW7
6	1465	36.4	1749	4	09H193
7	1465	36.4	1757	4	09H194
8	1463.5	36.3	1770	4	09H198
9	1463.5	36.3	1805	4	09H198
10	1412	33.0	1826	4	09H198
11	1396.5	34.7	1921	5	001349
12	1396.5	34.7	1921	5	09V7C9
13	1284.5	31.9	1595	5	020888
14	1174.5	29.1	1648	4	015058
15	1147	28.5	928	5	009997
16	1128	28.0	1174	5	09VB25

17	1076.5	26.7	1499	13	09DDA6
18	1035	25.7	1013	5	0960F4
19	1035	25.7	1121	5	09V1P4
20	1035	25.7	1121	5	016866
21	1034	25.7	1121	5	018390
22	1019.5	25.3	676	3	09C2M3
23	1004	24.9	429	4	09BQW5
24	978.5	24.3	1238	4	09HC12
25	847.5	21.0	752	4	09NKN9
26	807.5	20.0	1038	11	099PW8
27	799	19.8	744	13	093478
28	796	19.8	2756	10	09LJ50
29	783	19.4	699	13	098711
30	781.5	19.4	677	5	09VRK9
31	779.5	19.3	976	5	0969C2
32	775	19.2	1229	10	09LXV6
33	769.5	19.1	672	5	096V93
34	766	18.9	782	5	096V92
35	763	18.9	1099	5	09U142
36	761	18.9	782	5	019633
37	753.5	18.7	646	5	0965T6
38	750.5	18.6	784	5	0961H5
39	748	18.6	735	5	09U0D5
40	744.5	18.5	644	5	096V90
41	743.5	18.4	697	5	09U921
42	735.5	18.3	927	3	09HES9
43	735.5	18.3	1128	5	09U179
44	728.5	18.1	2158	10	09LUM5
45	722	17.9	929	3	P78718

## ALIGNMENTS

RESULT	ID	Query Match	Length	DB ID	Description
1	09V776	40.6	1773	5	09V776
2	09V776	40.5	1671	5	09V776
3	01-MAY-2000 (TREMREL. 13, last sequence update)	39.5	1581	5	018778
4	01-MAY-2000 (TREMREL. 13, last sequence update)	37.9	2205	5	09NGQ2
5	01-JUN-2001 (TREMREL. 17, last annotation update)	36.4	1749	11	09ECW7
6	CG8566 PROTEIN.	36.4	1749	4	09H193
7	UNC-104 OR CG8566.	36.4	1757	4	09H194
8	Drosophila melanogaster (Fruit fly).	36.3	1770	4	09H198
9	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	36.3	1805	4	09H198
10	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	33.0	1826	4	09H198
11	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	34.7	1921	5	001349
12	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	34.7	1921	5	09V7C9
13	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	31.9	1595	5	020888
14	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	29.1	1648	4	015058
15	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	28.5	928	5	009997
16	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:	28.0	1174	5	09VB25

RA Lasko P., Lelei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,  
 RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smilh T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirtsas R., Vektor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*.",  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003805; AAF57957.1; -.  
 DR HSSP; P17119; 3KAR.  
 DR FlyBase; FBgn0034155; unc-104.  
 DR InterPro; IPR000253; FHA\_domain.  
 DR InterPro; IPR001752; kinesin.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00498; FHA: 1.  
 DR Pfam; PF00225; kinesin: 1.  
 DR Pfam; PF00169; PH: 1.  
 DR PRINTS; PR00380; KINESINHEAVY.  
 DR SMART; SM00240; FHA: 1.  
 DR SMART; SM00129; KISC: 1.  
 DR SMART; SM00233; PH: 1.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1: 1.  
 DR PROSITE; PS00067; KINESIN\_MOTOR\_DOMAIN2: 1.  
 DR PROSITE; PS00003; PH\_DOMAIN: 1.  
 KW ATP-binding; Coiled coil; Microtubules; Motor protein.  
 KW SEQUENCE 1773 AA; 200755 MW; 335BE9C05E4DB691 CRC64;

Query Match	40.6%;	Score 1635;	DB 5;	Length 1773;
Best Local Similarity	44.5%;	Pred. No. 3.7e-86;		
Matches 367;	Conservative 143;	Mismatches 193;	Indels 122;	Gaps 20;

[illegible][illegible]

```

Oy      736  -----GTPRSDDDGDALFFGDKKSQODASNVVVEELRQOQAO 772
          : | | | : | : | : | : | : | : | : | : | : | : |
Db      754  RKMRYHOFTSLRDDLMGNATFL-----KEANATSVELKKKKVQFO 792

```

RESULT 2	
Q9NBL1	
ID Q9NBL1	PRELIMINARY; PRT; 1671 AA.

DT 01-OCT-2000 (TREMblrel. 15, created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE KINESIN SUPERFAMILY MEMBER DUNC104.

0S *Drosophila melanogaster* ('Fruit fly').  
0C *Eukaryota: Metazoa: Arthropoda: Tracheata: Insecta:*  
0C *Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:*  
0C *Ephydroidea: Drosophilidae: Drosophila.*  
0X NCBI\_TaxID=7227;

RA Grossberger R., Saxton W.M., Dickson B.J.; Dunc104."  
RT "Characterization of the Drosophila Unc104/KIF1A homolog,  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF247761; AAF74192.1; -  
DR HSSP: P17119; 3KAR.  
DR Flybase: FBGN0034155; unc-104.  
DR InterPro: IPR000253; FHA\_domain.  
DR InterPro: IPR001752; kinesin.  
DR InterPro: IPR001849; PH.  
DR Pfam: PF00498; FHA: 1.  
DR Pfam: PF00225; kinesin: 1.  
DR Pfam: PF00169; PH; 1.  
DR PRINTS: PR00380; KINESINHEAVY.  
DR SMART: SMO0240; FHA: 1.  
DR SMART: SMO0129; KISC: 1.  
DR SMART: SMO0233; PH; 1.  
DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1: 1.  
DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2: 1.  
DR PROSITE: PS50003; PH\_DOMAIN: 1.  
KW ATP-binding; Coiled coil; Microtubules; Motor protein.  
SQ SEQUENCE 1671 AA; 189430 MW; EAC3F1D3023360AF CRC64;

Query Match	40.58;	Score 1630.5;	DB 5;	Length 1671;
Best Local Similarity	44.38;	Pred. No. 6.2e-86;		
Matches 364;	Conservative 141;	Mismatches 188;	Indels 129;	Gaps 20

QY 66 NIKVVRVPFNAREIDRGAKCIVRNEGNOTLLPPEGAEBKARRSGATIMDGKAFAD 65  
::||| |||||:||| :|||: | :| :| :| :| :|  
Db 3 SVKAAVRVPFNSRELTARESKCIEMAGATTATNP-----KVPTSDSVKRRNFD 54

QY 66 RSYWSEDKNAIPNAROEIDLFGVPLLDNAEKGINNCIFAYGQTGSKSSTSMNGYKE- 12

Db 55 YSHHDDHDDDFSTQSNVYKDIGEEMLOHSDGYNVCIFAAGQGAACKSTMMRGQEQ 114  
 QY 125 -HGVIPIRCODMFRRINELQDKNLCTVEVSYLEIYNERVDLNPSTKGLKREHPS 183  
 Db 115 QEGITPMICKDLFRIOPTERD-DLKYSEVSYMEIYCERDRLNPKNGKLARKREHPL 173  
 QY 184 TGPVYEDLAKLVRSFOEIEMLDEGNKARTVAATNMNETSSRSNAVFTLTJQRMDEE 243  
 Db 174 LGPVYEDLSKLAVTDYQDIHDLIDEGNKARTVAATNMNETSSRSNAVFTLTJQRMDEE 233  
 QY 244 TKMOTERAKISLVDLASERATSTGATGARKLKEGAEINRSLSITGRIYALADMSGKQ 303  
 Db 234 TNLTEKYSKISLVDLASERADSTGATGARKLKEGAMNKSITLTGKYISALAEVSAKSK 293  
 QY 304 ---KNOLVPRDSVLTWLLKDSLGNSMTAMIAISPADINFEETLSLRADGAKRIK 360  
 Db 294 KNTKKADIPYRDSALTYLLRENLOGNSKTAMIAISPADINDETSLTRADAKOIV 353  
 QY 361 NHAUVNEDPNAMIRELKEELAQLRSKLQSSGGGGAGSGGPVEESYPPDTPLEKOIV 420  
 Db 354 CKAUVNEDANAKLIRELKEEIOKLELDLKAEG-----I 386  
 QY 421 SIQOPD-----ATYKKMSKAE-----IYEOLNOSKLYRDLNQTWEKLTATEBH 466  
 Db 387 EVQEDDELTKSTVLSKPSKSRNRNGSTTEMAVDQLOASEKLAELNETWEKLTATEIR 446  
 QY 467 KEREALEELISL-KEGF-VGPIYKEMPHLVNLSDDPLLAECVYNIKPGOTRGNVN 524  
 Db 447 VQREAVFAMKGVAVKEDITGVFSPKTPHLVNLNEDPNLSECLLYIKELTFLGHE 506  
 QY 535 QDTQAEIRLNGSKLKEKCTEENDVNTIYVNEKAAVNCGVRIKPTRLNSGRIILG 584  
 Db 507 ANVPDDIOLSSHIKEKCTEENKSTYTLPLPKDAIIVYNGRKLYEEVLTGSRVILG 566  
 QY 585 DEHIFREHPEAREBEREQSLRHSVTNSQLGSPAPGRHDTLSKASDADGDSRSP 644  
 Db 567 KNHVFRTPNPDARE-----LRDKT-----ETENEAEVEV--KTD- 601  
 QY 645 LPHFGKDSDFYARRENAASITLGD-----QKISHLTDD-----ELDALFDDVOK-- 690  
 Db 602 -----QOVNMFACCELLEK-QGIDLKAEMKKRLDNLEQYKRRKLDADQDFEQRTY 654  
 QY 691 -AR-AVRGLVEDNEEDSOSSPYRKDYNSGTIDNLSLDTAITMP----- 735  
 Db 655 EARIDALOKOVEOSMTSMYSYSPEDFHOEDYTTNPMESCTAREAGIAANAFKRW 714  
 QY 736 -----GTPRSDDDALFPFGDKSKODASNDVEELRQOQAG 772  
 Db 715 RYHOFTSLRDLDMGNATFL-----KEANAISVELLKKRVQFO 750  
 RESULT 3  
 ID 018778 PRELIMINARY; PRT: 1581 AA.  
 AC 018778;  
 DT 01-NOV-1996 (TREMBLREL 01, Created)  
 DT 01-DEC-2001 (TREMBLREL 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLREL 19, Last annotation update)  
 DE HYPOTHEICAL 179.4 KDA PROTEIN.  
 GN C52E12.2  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE-99069613; PubMed-9851916;  
 RA None;  
 RT \*Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.\*;  
 RL Science 282:2012-2018(1998).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Du 2.?  
 RT "The sequence of C. elegans cosmid C52E12."  
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U50135; AA93453.2; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1581 AA; 179408 MW; 61CA9CA0B322CCCC CRC64;

Query Match 39.5%; Score 1590.5; DB 5; Length 1581;  
 Best Local Similarity 44.7%; Pred. No. 1.2e-83;  
 Matches 356; Conservative 139; Mismatches 213; Indels 89; Gaps 19;

QY 6 NIKVYVRPPEARELIDGAKCIYMEGNOTILTPPGAEEKARKSGKINDGPKAFAD 65  
 Db 3 SVKAVARVPFQOREISNTSKCVLDVNGMTTIT-----NGHSINRENSFND 50  
 QY 66 RSYWSEDKNAPYARQEDLFODLGVPLDNMAFKYNNCIFAYGOTGSKSGSMGYGK-- 123  
 Db 51 HSYWSEFARNDFPFIQKQYVELGYEMLEHAFEGYNVCIFAAGQGAACKSTMMRGKANDP 110  
 QY 124 -EHGVIPIRCODMFRRINELQDKNLCTVEVSYLEIYNERVDLNPSTKGLKREH 182  
 Db 111 DMGIIPRLCNDLFARIDN-NNDKVOYSEVSYMEIYCERKDLNPNSGNGLARKREH 169  
 QY 183 STGPVYEDLAKLVRSFOEIEMLDEGNKARTVAATNMNETSSRSNAVFTLTJQRMDE 242  
 Db 170 ILGPVYDOLTKAACSNDICMLMBGNAKARTVAATNMNETSSRSNAVFTLTJQRMDE 229  
 QY 243 ETKMDTEKVAKISLVDLASERATSTGATGARKLKEGAEINRSLSITGRIYALADMSGK 302  
 Db 230 DSNLDTKHSKISLVDLASERANSTGAGQRLKEGANINKSLTTLGLVYSKLAESTBK 289  
 QY 303 OKKNO-LVPIYRSVLTWLLKDSLGNSMTAMIAISPADINFEETLSLRADAKRIK 361  
 Db 290 KSNKGVPIYRDSVLTWLLRENLOGNSKTAMIAISPADINDETSLTRADAKOIVC 349  
 QY 362 HAVVNEDPNAMIRELKEELAQLRSKLQSSGGGGAGSGGPVEESYPPDTPLEKOIVS 421  
 Db 350 QAVVNEDPNAKLIRELNEEVITLRHLKDG-----IDVTD 385  
 QY 422 IQQPDATYKKMSKA-----EIVEOLNOSKLYRDLNQTWEKLTATEIRKREALAEEL 476  
 Db 386 VQETPGKHKKKPKLPAHVHEQLEKLOESEKLAELIGKTEQGLIHTEIRKREELRBM 445  
 QY 477 GIS-IEKG-FVGPIYKEMPHLVNLSDDPLLAECVYNIKPGOTRGNVNOTQAEIRLN 534  
 Db 446 GLACAEDEGTTGLVFPKPKLPHLVNLNEDPNLSECLLYIKELVTSGRVAVHHRDILS 505  
 QY 535 GSKIRKEHCTEENDVNTIYVNEKAAVNCGVRIKPTRLNSGRIILGDEHIFREHNP 594  
 Db 506 GAILLELHCEPINEGANTYLTMKPNASCTINKOVTTTPVLTGSRVILGHEHNVFRYND 565  
 QY 595 EBAREREOGSLRHSVTNSQLGSPAPGRHDTLSKAGSDADGDSRSDPLPHFGKDS 654  
 Db 566 QBARSRHNLAIARQPIDMKYA-----QOELDKOGIDLKADMEK-----KMLEMESQ 614  
 QY 655 WFIARREASALIGDOKISHTDELALFPDVOKARAVR-----GLVEDNED--SD 706  
 Db 615 Y---RREKVE---LEQKYHOT-REYESMIENLOKOVLAOSYISGGSGIMEGRMLTS 666  
 QY 707 SOSSEFPVRDKYNSGTIDNLSLDTAITMP-----GTPRSDDDDALFPFGDKSKODASND 762  
 Db 667 SLLEPPEELKWTSDQ--KRVYLKAATIKRYHOFTSVYRDDLMGNATFLV-----KEANAIS 718

```
QY      763 VEELRQQAQMEALKT 779
      || :: | : || |
Db      719 VE--LKKKVQGFALLT 733
```

RESULT	4		
09NGC02		PRELIMINARY;	PRT: 2205 AA.
ID	09NGC02		
AC	09NGC02;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	KINESIN UNC104/KIF1A HOMOLOG.		
GN	UNC104.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.		
CC	NCBI_TaxID=44689;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20014990; PubMed-10545495;		
RA	Pollack N., de Hostos E.L., Tuck C.W., Vale R.D.;		
RT	"Reconstitution of membrane transport powered by a novel dimeric		
RT	kinesin motor of the unc104/Kif1a family purified from		
RT	dictyostelium.";		
RL	J. Cell Biol. 147:493-506(1999).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RA	Pollack N., Vale R.D.;		
RL	Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.		
DR	HSSB; AF245277; AA63384.1; '-'		
DR	EMBL; P17119; 3KAR.		
DR	InterPro: IPR001752; Kinsln.		
DR	InterPro: IPR001849; PH.		
DR	Pfam; PF00225; kinesin_1.		
DR	Pfam; PF00169; PH; 1.		
DR	PRINTS; PR00380; KINESINHEAVY.		
DR	SMART; SM00129; KISC; 1.		
DR	SMART; SM00233; PH; 1.		
DR	PROSITE; PS00411; KINSIN MOTOR DOMAIN1; 1.		
DR	PROSITE; PSS0067; KINSIN MOTOR DOMAIN2; 1.		
DR	PROSITE; PSS0003; PH_DOMAIN; 1.		
DR	ATP-binding; Coiled coil; Microtubules; Motor protein.		
SW	SEQUENCE 2205 AA: 248002 MW: 02C5101E9D61C9ED CRC64;		

Query Match	37.9%	Score 1525.5	DB 5	Length 2205
Best Local Similarity	45.1%	Pred. No. 1.1e-79		
Matches 333	Conservative 118	Mismatches 167	Indels 121	Gaps 13

QY	6	NKVVVVRPFPNARIBIDGACIYIMEMNOQIILPPGAEKARKSGKTIIMDPKPAFAFD	65
Db	2	NVQVAVRVRPFSREKERNAEIYOMNKKSTILRPSAL--RANPLAAPADDEKSPSFD	59
QY	66	RSYMSFDKNARPYAQEQLFDLDGVPLLDNAFKCIYNNCIRAYGOTGSGKSYSMWGCKEH	125
Db	60	YSYMSYSDNDPHEFASQSYVYNDILKEVYLKKNMDFNCSIFAYGGTGSGKSYSMWGGEK	119
QY	126	GYIPRICDMFRINELDKDN--LTCIVEYSLEYIENERYRDLNBSK--GNLKYREH	181
Db	120	GIILPICELNQRIOISTPNSNEQIYKTYVSWMEIYNEKUKDOLLNPNKKTGSLKAYRNN	179
QY	182	PSTGCVYEDLAKLVYRSQOEIENLMDGKNKRITYAALNNMETSSRSYAVFTLLITQKMH	241
Db	180	PSTGCVYEDLSKLAVKSPSEIDMLMDEBSKARTYASTMNTSSRSYAVFTIYFTQKID	239
QY	242	EETKMDTEKVAIKSYVDLAGSERATSTGATCARLKEGAELIRSLSTGRVYLAALADSSG	301
Db	240	KTRGTAIDRVSXISLVDLAGSERANSTGATGVRLKEGANINKSLSTLGKVIYSLAEVST-	298
QY	302	KOKNOVLVPRDSVLTWLKDSLGNSKMTAMIAISPADINFEETLSTLRYSAKRIKN	361
Db	299	SKKAVFVPRDSVLTLYLLETIGNSKPTIMIAISPADINVEESLSTLRYSAKRIKT	357

```

QY 362 HAYVNEDPMAMTELKEEELAAQLESLQSSGSGGGGAGSGSGPYEESYPRDFLEMQIYS 421
    |||||:::||||:::||||:::|
Db 358 VAVVNEDQOSKLTELQGEVERLEPMAADQGS-----QHANDSKLM--- 399
QY 422 IQOPDATYKKKSKAIEYVOLNOSKELRYDLNQWTEERLAKTEELIHKREKRAALELGISIE 481
    ||:::||||:::||||:::||||:::|
Db 400 -SDYDELYSTLN-----EKIQYERKMAELNKSWEKLESEALRREDMAALMDMGAIK 453
QY 482 KGFVGRYHKSMPHLVNLSDPPLAEBCLVYNIRFGQTRVGNVNDTOAEIRLNGSKLTKE 5411
    |||||:::||||:::||||:::|
Db 454 V-----VSSIPHLINELNEDPLMSESLIYYKKGKTRIGRSDSEIPQDIIILGNLNIKE 506
QY 542 HCTFENDVNVTYIYPN----- 557
    |||||:::|:|
Db 507 HCIFENINGKVIISPSPFNMMNNNNKNSSTPTSSKSPKPSKEKEKENNDDDGEEK 566
QY 558 --EKAAYVNVCVRIDKPTRLRSGRIILGDHIEFRFNHPEARARQEOSILRHVSYSNQ 615
    :::|||||:::|:|:::|||||:::|:::|
Db 567 KLDRSYIYVNGVEINKPPIITLTGRVILIGNNHIFRFPNPEBALIABERKQTTGGIYSS- 625
QY 616 LGSPPAGRHDRITLSKAGSDADGDSRSDSPLRHFGKSDSWEFYARREASAAILGIDOKISH 675
    :::|:::|:|
Db 626 -----TKSPVQOI-----MDYDFALNELAS-IQGTLSAKSH 655
QY 676 LTD-----DELDALFDVQ 689
    :::|:::|
Db 656 INDKOYKKOMRALYDQIR 674

```

```

RESULT      5
Q9E0W7      PRELIMINARY;      PRT;      1749 AA.
AC      Q9E0W7;
DI      01-MAR-2001 (TREMBLrel. 16, Created)
DI      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DI      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      KIF13A.
GN      KIF13A OR KIF13A.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ICR; TISSUE=BRAIN;
RX      MEDLINE=20560742; Pubmed=11106728;
RA      Nakagawa T.; Setou M.; Seog D.H.; Ogasawara K.; Dohmae N.; Takio K.;
RA      Hirokawa N.;
RT      "A Novel Motor, KIF13A, Transports Mannose-6-Phosphate Receptor to
RT      Plasma Membrane through Direct Interaction with Ap-1 Complex.";
RL      EMBL; AB037923; BAB16346.1; -.
DR      HSSP; P33176; 1BG2.
DR      MGD; MGI:1098264; Kif13a.
DR      InterPro; IPR000253; FHA_domain.
DR      InterPro; IPR001752; Kinesin.
DR      Pfam; PF00225; Kinesin; 1.
DR      PRINTS; PR00380; KINESINHEAVY.
DR      SMART; SM00240; FHA; 1.
DR      SMART; SM00129; KISC; 1.
DR      PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW      ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ      SEQUENCE      1749 AA;      195812 MW;      DA9E026127DDBF2 CRC64;

```

Query Match	36.4%;	Score 1468;	DB 11;	Length 1749;
Best Local Similarity	44.0%;	Pred. No. 1.8e-76;		
Matches 342;	Conservative 112;	Mismatches 223;		
			Indels 100;	Gaps 15

```

Oy      7 IKVVRVRPFPNAKEIDRGAKCIYMEGNOTILTPPGAEEKARKSGKTMIDGPKAFAPDR 66
        :: ||||| | :: |:: |||||: ||| :: | |||||
Db      6 VKVAVRVRPMPNRRELELNTKCVAMEGNGQVLPHPSPNTKOGERK-----PPKVFAPEDY 59

```



AC Q9H194;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE KINESIN-13A1.  
GN KIN13A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Jamin S., Quach H., Fellous M., Bourgeron T.;  
RT "Identification of the human KINESIN-13a gene, homologous to  
RT Drosophila kinesin-73 and candidate for schizophrenia";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ291578; CAC20442.1; -.  
DR HSSP; P33176; 1862.  
DR InterPro; IPR000253; FHA\_domain.  
DR Pfam; PF00225; kinesin.1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00240; FHA.1.  
DR SMART; SM00129; KISC.1.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2.1.  
DR ATP-binding; Coiled coil; Microtubules; Motor protein.  
KW SEQUENCE 1757 AA; 197084 MW; 92277256f8d8e0f CRC64;

Query Match 36.4%; Score 1465; DB 4; Length 1757;  
Best Local Similarity 43.6%; Pred. No. 2.7e-76;  
Matches 343; Conservative 119; Mismatches 205; Indels 120; Gaps 19;

QY 7 IKVVRVPFPAAREIDRGAACIVRMEGNOTILTPPGAEEKARKSGKTIIMGPKAFADR 66  
DB 6 KVAVRVPFPAAREIDRGAACIVRMEGNOTILTPPGAEEKARKSGKTIIMGPKAFADR 59  
QY 67 SYWSEDK NARNYARQEDLPDGLVPLIDNAFKGINNCIFAYGOTSGSKYSMMGYGKEH 125  
DB 60 CFWSMDESNTTKYAGQEVFVFCLEGLIEKAFQGYNCIFAYGOTSGSKYSMMGYGKEH 119  
QY 126 GVPRICODMFRINELQKDNLCYEVSVLEIYNEVRDLNP-STKGINKVAREHST 184  
DB 120 GLIPRLCALFKRIS-LEQNSQTFKVEVSWEIYNEVRDLNP-STKGINKVAREHST 178  
QY 185 GPYEDLAKLVVRSFOELENLMDGKARKATVAATNMNETSSRSHAVFTLTQKHDEET 244  
DB 179 GPYDGLSQLAVTSPEDIESLMSGKSRVYAAATNMNETSSRSHAVFTLTQKHDEET 238  
QY 245 KMDTEKAKISLVNLAGSERATSTGATGARLKEGAELNRSLSLTGRVIAALADSSGKQK 304  
DB 239 GNSGEKSKVSLVNDLAGSERATSTGATGARLKEGAELNRSLSLTGRVIAALADSSGKQK 297  
QY 305 KNOIVPYDVLVTLTKSLGNSMTAMTAAISPADINFEETISLTGRVADAKRKNVAV 364  
DB 298 KSKVPPYDVLVTLTKSLGNSMTAMTAAISPADINFEETISLTGRVADAKRKNVAV 357  
QY 365 VNEDPNAMIRELKEELAQLRSKLSQSGGGGAGGSGGPEESYPPDTPEKQIVSIQO 424  
DB 358 VNEDPNAMIRELKEELAQLRSKLSQSGGGGAGGSGGPEESYPPDTPEKQIVSIQO 384  
QY 425 PDATVKKSKRAEIVQNLQSKLYRDLNQTWEKLAKEEIHKEREALBELGISIEKGF 484  
DB 385 -----EAKKAKELKEEIESEKELIKELVTWEKLRKTEEIAERQRLSESGISLEKS- 438  
QY 485 VGPYHSKEMPLVNLSDPLLAEGLYVNIKRGQTRVGVNVDQAEITLNSKILKEKCT 544  
DB 439 -GIIVGDDKCTLVNLNDPALNELLYTLK-DHTRVG---ADTSODIQLFGIGQPOHCE 493  
QY 545 FEVY-DNVTIVIPNEKAIVAVNGVARDKPTLRSGYRIITLDFHIFRNHEEAREAEQ 603  
DB 494 IDIASDGDVTLTPKRNEMASCYNGTIVCGSTIQLMHDRILMGNNHFRILNKKRRRWMLK 553

QY 604 QSLRHSVTNSQSGPAPGRDRTLKAGSDAGDSRSDSPLPHPRKGSQMPYARREA 663  
DB 554 D-----FEKETGPP-----EHD-----LDAASEASSE-----PDYNEFEAKMEVI 588  
QY 664 SAILGLDOKISHLTDDELDAFPDQKARAVRGRGVEINE-----SDSOGSSFPV 713  
DB 569 MKTINSNDPQNVN-QVEKQYLEKRSALREQLMYRELEQLRQULSPROPQSSGPD 647  
QY 714 RDKYMSNGTIDNFSIDTAITMGTPRSSDDGDALFFGDKSKODASNYDVELRQQAQM 773  
DB 648 RLAYS-----QTA-----OQKTYMAEERD-ELFRQSLAKL 678  
QY 774 BEALKTA 780  
DB 679 REQLVKA 685  
RESULT 8  
Q9H186 PRELIMINARY; PRT; 1770 AA.  
AC Q9H186;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE KINESIN-LIKE PROTEIN RBKIN2.  
GN RBKIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RA Chen D., Squire J.A., Gallie B.L.;  
RT "Genomic amplification in retinoblastoma narrowed to 1.2 Mb on  
RT chromosome 6p containing a novel kinesin-like gene, RBKIN";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY014404; AAC38891.1; -.  
DR HSSP; P33176; 1862.  
DR InterPro; IPR000253; FHA\_domain.  
DR InterPro; IPR001752; kinesin.  
DR Pfam; PF00225; kinesin.1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00240; FHA.1.  
DR SMART; SM00129; KISC.1.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2.1.  
DR ATP-binding; Coiled coil; Microtubules; Motor protein.  
KW SEQUENCE 1770 AA; 198550 MW; D926CDD8ED8CB3IC CRC64;

Query Match 36.3%; Score 1463.5; DB 4; Length 1770;  
Best Local Similarity 45.5%; Pred. No. 3.3e-76;  
Matches 330; Conservative 116; Mismatches 189; Indels 91; Gaps 16;

QY 7 IKVVRVPFPAAREIDRGAACIVRMEGNOTILTPPGAEEKARKSGKTIIMGPKAFADR 66  
DB 6 KVAVRVPFPAAREIDRGAACIVRMEGNOTILTPPGAEEKARKSGKTIIMGPKAFADR 59  
QY 67 SYWSEDK NARNYARQEDLPDGLVPLIDNAFKGINNCIFAYGOTSGSKYSMMGYGKEH 125  
DB 60 CFWSMDESNTTKYAGQEVFVFCLEGLIEKAFQGYNCIFAYGOTSGSKYSMMGYGKEH 119  
QY 126 GVPRICODMFRINELQKDNLCYEVSVLEIYNEVRDLNP-STKGINKVAREHST 184  
DB 120 GLIPRLCALFKRIS-LEQNSQTFKVEVSWEIYNEVRDLNP-STKGINKVAREHST 178  
QY 185 GPYEDLAKLVVRSFOELENLMDGKARKATVAATNMNETSSRSHAVFTLTQKHDEET 244  
DB 179 GPYDGLSQLAVTSPEDIESLMSGKSRVYAAATNMNETSSRSHAVFTLTQKHDEET 238  
QY 245 KMDTEKAKISLVNLAGSERATSTGATGARLKEGAELNRSLSLTGRVIAALADSSGKQK 304



Db 239 GNSGKSVKSVSLVDLAGSERVSKTGACGERLKEGNSINKSLTTGLVTSIADQAAGK-G 297

Qy 305 KNOIYPYRDSVLTWTLKDSLGSNSMTAMIAISPADINEETLSTLRVADSARKIKNHAV 364

Db 298 KSKFPYRDSVLTWTLKDNLGNSTSMIATISPADNTEETLSTLRADRAKRIYNHAY 357

Qy 355 VNEEDNARMIRELKEELAQLSKLSQSGGGGAGCGGPGVEESYPPDTPLEKQIVS10Q 424

Db 358 VNEEDNARMIRELKEELKEELQSLQA----- 384

Qy 425 PDATYKKMSKAIYVQLONSEKLYRDNLQNTWEKLAETEEIHKEREALAEKISIEKGF 484

Db 385 -----EAKAPLKEKELESEKELIKELVTWEKLRKTEELAEQERKQLESIGISLEMS- 438

Qy 485 VGPYHSEKMPHLVNLSDPLAECLVYNIKPQTRVGNVNDTOAIEIRLNGSKILKEHCT 544

Db 439 -GIRYGDCKCYLVNLMNDPALNELLYLYLK-DHTRVG---ADTSODIQLFQIGIQPQHC 493

Qy 545 FEWV-DNVTYIVPNEKAAMVNGVRIDKPTRLSGYRIILDFHIFRPNHPEAPAEOR 603

Db 494 IDIASDGVTLTPKRNARSCVNGTLVCTQLMHGDRIILMGNNHFRILNPKRRDMLK 553

Qy 604 QSLRHSVTNSQLGSPAPGRDRTLSKAGSDADGSRSDSPLPHRGKDSMFVAREEA 663

Db 554 D-----FEKETGPP---EHD-----LDASEASSE-----PDYNTFEFQMEYI 588

Qy 664 SAILGLDQKISHLTDELDAFLPDVQKARAVRGIVEONE-----DSDSQSSFPV 713

Db 589 MKTLSNPNVQNVV-QVLEKQYLEEKRSALAEQRLMYERELEQLRQQLSPDQPOSSGPD 647

Qy 714 RDKYMS 719

Db 648 RLAYSS 653

RESULT 9

Q9NH9 PRELIMINARY; PRT; 1805 AA.

AC Q9NH9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE KINESIN-LIKE PROTEIN RAKIN1.

CN RAKIN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=RETINA;

RA Chen D., Squire J.A., Gallie B.L.;

RT \*Genomic amplification in retinoblastoma narrowed to 1.2 Mb on chromosome 6p containing a novel kinesin-like gene, RAKIN.1;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AI014403; AAC38890.1; -

DR HSSP; P33176; 1BG2.

DR InterPro; IPR000253; FHA\_domain.

DR InterPro; IPR001752; kinesin.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PRO0380; KINESINHEAVY.

DR SMART; SM00240; FHA; 1.

DR SMART; SM00129; Kisc; 1.

DR PROSITE; PSS0067; KINESIN\_MOTOR\_DOMAIN2; 1.

DR ATP-binding; Colled coil; Microtubules; Motor protein.

SO SEQUENCE 1805 AA; 202258 MW; AAF249B859A04B7A CRC64;

Query Match 36.3%; Score 1463.5; DB 4; Length 1805;

Best Local Similarity 45.5%; Pred. No. 3.4e-76;

Matches 330; Conservative 116; Mismatches 189; Indels 91; Gaps 16;

Qy 7 IKVVAVRVPFNAIREIDRAKCIIVRMENQNTILTPPGAEKARKSGKTIIMGPKAFADR 66

Db :|| ||||| | ||: ||:| ||||| | ||| :| |||||

Qy 6 VQVAVRVPFNAIRELELMTKCYVEMEGNQTIVLHPPSMTKQGERK-----PPKVFADY 59

Qy 67 SYWSDK-NAPYARQEDLFODLGVPLDNARKYNNCCFAFGQSGSGSYMMQYGENH 125

Db 60 CFWSDSTWTKYAGQEVVFKCLGRGLEKAFQGNACIPAFQGTGSGSFSMGMHADL 119

Qy 126 GVIPRICODMPFRINELQDKMTCTVEYSYLEIYNERVRLNLP-STFGKNGLVHEPST 184

Db 120 GLIPRLCCALFKRIS-LEQNESQTFKVEYSYEITNEYRDLDDPKSGNSQSLKVEHYL 178

Qy 165 GPYEDLAKLVVRSPOEILNIDEGNKARTVAATNNETSSSHAVFTLTLOKHDEET 244

Db 179 GPYVGLSOLAVTSFEDIESLMSSEGNKSTVAATNMNESSRSHAVFNITITQTYLQDS 238

Qy 245 KMDTEKVKISLYDLAGSERATSGATGARKKEGAENRSTSLGRVLAALADMSGKQ 304

Db 239 GNSGKSVKSVSLVDLAGSERVSKTGACGERLKEGNSINKSLTTGLVTSIADQAAGK-G 297

Qy 305 KNOIYPYRDSVLTWTLKDSLGSNSMTAMIAISPADINEETLSTLRVADSARKIKNHAV 364

Db 298 KSKFPYRDSVLTWTLKDNLGNSTSMIATISPADNTEETLSTLRADRAKRIYNHAY 357

Qy 355 VNEEDNARMIRELKEELAQLSKLSQSGGGGAGCGGPGVEESYPPDTPLEKQIVS10Q 424

Db 358 VNEEDNARMIRELKEELKEELQSLQA----- 384

Qy 425 PDATYKKMSKAIYVQLONSEKLYRDNLQNTWEKLAETEEIHKEREALAEKISIEKGF 484

Db 385 -----EAKAPLKEKELESEKELIKELVTWEKLRKTEELAEQERKQLESIGISLEMS- 438

Qy 485 VGPYHSEKMPHLVNLSDPLAECLVYNIKPQTRVGNVNDTOAIEIRLNGSKILKEHCT 544

Db 439 -GIRYGDCKCYLVNLMNDPALNELLYLYLK-DHTRVG---ADTSODIQLFQIGIQPQHC 493

Qy 545 FEWV-DNVTYIVPNEKAAMVNGVRIDKPTRLSGYRIILDFHIFRPNHPEAPAEOR 603

Db 494 IDIASDGVTLTPKRNARSCVNGTLVCTQLMHGDRIILMGNNHFRILNPKRRDMLK 553

Qy 604 QSLRHSVTNSQLGSPAPGRDRTLSKAGSDADGSRSDSPLPHRGKDSMFVAREEA 663

Db 554 D-----FEKETGPP---EHD-----LDASEASSE-----PDYNTFEFQMEYI 588

Qy 664 SAILGLDQKISHLTDELDAFLPDVQKARAVRGIVEONE-----DSDSQSSFPV 713

Db 589 MKTLSNPNVQNVV-QVLEKQYLEEKRSALAEQRLMYERELEQLRQQLSPDQPOSSGPD 647

Qy 714 RDKYMS 719

Db 648 RLAYSS 653

RESULT 10

Q9NQT8 PRELIMINARY; PRT; 1826 AA.

AC Q9NQT8;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE KINESIN-LIKE PROTEIN GAKIN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=20435838; PubMed=10859302;

RA Hanada T., Lin L., Tibaldi E.V., Reinherz E.L., Chishti A.H.;

RT \*GAKIN: a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes.;

RL J. Biol. Chem. 275:28774-28784(2000).

DR EMBL; AF279865; AAF81263.1; -



QY 485 VGPYHSEKMPHLYNLSDDPLAECLYVNIKPGQTRVGNVNDQTAELRNGSKILKEHCT 544  
 Db 439 -GIKYEKNKYYLVNLMADSLNELLYYLKRTLLIGKRTISQOQPIQJLSGLGIOPHCY 497  
 QY 545 FENVNDVVTIVPNEKAAMVNGVRIDKPTRLRSRYRIILGDFHIFRNFHPE----- 595  
 Db 498 ITIEDSGILMEYVGARCFVNSAAVEKTPPLQNGDRILLGNHFFRVNPSKNNTSMCAS 557  
 QY 596 --EARAEROQSLRHSVTNSOLGS 618  
 Db 558 EPOTPAQLIDYNFARDEIMQNELSN 582  
 RESULT 12  
 Q9V7C9 PRELIMINARY; PRT; 1921 AA.  
 ID 09V7C9;  
 AC 09V7C9;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE KHC-73 PROTEIN.  
 GN KHC-73 OR CG8183.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.F., Baas A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Doudon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwan C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson S.M., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT \*The genome sequence of Drosophila melanogaster.\*  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003810; AAF58129.1; -  
 DR HSSP; P33176; 1B62.  
 DR FLYbase; FBgn0019568; KHC-73.  
 DR InterPro; IPR000938; CAP-Gly.

DR InterPro; IPR000253; FHA\_domain.  
 DR InterPro; IPR001752; kinesin.  
 DR InterPro; IPR001220; Lectin\_legB.  
 DR Pfam; PF01302; CAP\_GLY; 1.  
 DR Pfam; PF00225; kinesin; 1.  
 DR PRINTS; PR00380; KINESINHEAVY.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00129; KISC; 1.  
 DR PROSITE; PS00845; CAP\_GLY; 1.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS00657; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR PROSITE; PS00307; LECTIN\_LEGUME\_BETA; UNKNOWN\_1.  
 DR App-binding; Coiled coil; Microtubules; Motor protein.  
 KW SEQUENCE 1921 AA; 214997 MW; DEDCA75A6344C42 CRC64;  
 SQ  
 Query Match 34.7%; Score 1396.5; DB 5; Length 1921;  
 Best Local Similarity 48.8%; Pred. No. 2,8e-72;  
 Matches 305; Conservative 88; Mismatches 171; Indels 61; Gaps 8;  
 QY 7 IKVYRVPRFPNAREIDRGAKCIVRMEGNQITLTPPGAEEKARKSGKRTIMDGPKAFAPDR 66  
 Db 6 IKVAVRVPRFPNAREIDRGAKCIVRMEGNQITLTPPGAEEKARKSGKRTIMDGPKAFAPDR 58  
 QY 67 SYWSPDKNAPYARQEDLFQDLGVPLDPAFPGYNNCIFYAQGTGSGKSYSMGYGKENG 126  
 Db 59 CEYSINPEDEENFASQETVEDCVGRGILDNAFQYNAICIFAYGTGSGKSYTMWGQESKG 118  
 QY 127 VIPRICQDMERRINLEQKKNLTCTVEVSYLTIVYERVDLNPSTKNTLVREHPSRG 185  
 Db 119 IIPRLCDLFSAIAN-KSPPELMTYKEVSYMEIYVKNVLDLDPKNNKSLVREHNVG 177  
 QY 186 PVYEDLALVVSFOEITENDEGNKARTVAATNNMETSSRSNAVFTLTQKMHDEETK 245  
 Db 178 PLYDGLSQLAVSYXDIDMLNTEGKMSRTVAATNNMAESSRSHAVSVTLQILDQANG 237  
 QY 246 MDTEVYAKISLDVLAGSERATSTGATGARLKEGAEINSLSTLGRVIALADMSGQKQK 305  
 Db 238 VGEKYSRMSLDVLAGSERAVTGAAGVDRLEKGSINKSLTLGLVYSKLADQSNKQK 297  
 QY 306 N-OLVPRYSVLTWILKDSLGSNSMTAMTAISPADINEFTLSTLRVDSKRIKNH 364  
 Db 298 NKEFVPRYSVLTWILKDLNMGNSRTVMATISPSADNTEFTLSTLRVDRKRIYNH 357  
 QY 365 VVEDNPARIRLEKEBLQLRSKLQSGGGGAGGSGGSPVPPDPLEKQIYSIQ 424  
 Db 358 VVEDNPARIRLEKEBLQLRSKLQSGGGGAGGSGGSPVPPDPLEKQIYSIQ 389  
 QY 425 PPAVKKMSKADIVQNLQSEKLYRDNLQNTWEKLAETEEIKERBALEELGISTEKG 484  
 Db 390 -----GDVQKLAESENIMKOISQTEWEKLVKTERIONEROALAEKKGISVOAS- 438  
 QY 485 VGPYHSEKMPHLYNLSDDPLAECLYVNIKPGQTRVGNVNDQTAELRNGSKILKEHCT 544  
 Db 439 -GIKYEKNKYYLVNLMADSLNELLYYLKRTLLIGKRTISQOQPIQJLSGLGIOPHCY 497  
 QY 545 FENVNDVVTIVPNEKAAMVNGVRIDKPTRLRSRYRIILGDFHIFRNFHPE----- 595  
 Db 498 ITIEDSGILMEYVGARCFVNSAAVEKTPPLQNGDRILLGNHFFRVNPSKNNTSMCAS 557  
 QY 596 --EARAEROQSLRHSVTNSOLGS 618  
 Db 558 EPOTPAQLIDYNFARDEIMQNELSN 582  
 RESULT 13  
 Q20888 PRELIMINARY; PRT; 1595 AA.  
 AC Q20888;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE HYPOTHETICAL 179.4 KDA PROTEIN.



Qy	7	IKVAVRVRPFAPRAIDNGACCIYRMAGNQTILTRPGAEAKARKSGKITMDGPKAFADR	66
Db	359	VTAVRVRPFPRKREIKKQAQVYVMSGKETVYEH-----DTKOYTN---FYIDV	405
Qy	67	SYMSFEDKAPRYAQAQEDLPFDGLVPLLDNAFKYNNCIPAYGOTSGSKSYMAGKKEHG	126
Db	406	SPMSFEDCHPRYASQTLYEYKLAAPLLEAFREFBNCLTRLYQOTSGSKSYMAGSSEBPG	465
Qy	127	VIPRICDMERRINELQKDKNLCTVEVSYLEYNERVDLL---NESTGKLNKVBH	181
Db	466	IIPFCEDELSQVARKQ--TQEVSYHTIEMSEFFEYVYNEKIDHLLVCKDENQGRQPLRVREH	524
Qy	182	PSPTGPEYEDLAKLVRSFQSEIMLMQEGKAKRPVATNNENSSRSHAFVLTLLQKMD	241
Db	525	PVPTPEYEAISNIVSYSTADIQSWLEGMKNQRTATYAGNDMSRSHSVFTLYVMQYOTE	584
Qy	242	--EETKMDTEYKAKISLVDLAGSERATSTGATGARLKEGAELNRSLSYGRVYALADMS	299
Db	585	FVEGEEDHRTITSRINLIDLAGSERSTAHNTNDRLKEGVSYINKSLTLGKVTALSQDA	644
Qy	300	SGARKKQQLPYRDSVLTWLLKDSLGNSMTAMIAAISPADINFEETSLTLYADSAKRI	359
Db	645	N---QRSVFPIPRRSYVLTWLLKESLGNSMTAMIAAISPAISNIEETSLTLYAQARLI	701
Qy	360	KNAHVVEDNAPRARIRELKFEELQKLSKLOSSGGGGGAGSGGQVPEESYPRDPFLKQI	419
Db	702	VNIKAVVEDNAKILIRELKFEIKLKLAAQRNS-----	733
Qy	420	VSIOQDPATVYKMSKAETVE---QLNQSEKLYDLNQWTEERKLAKTEIHKEREALDEL	476
Db	734	---RNIDPERYRLCROEITSLRMLKHQERDMAEMQVWKKEFEQAEKRIKLETKELQKA	790
Qy	477	GISTEKFQVPRYSHKEMPHVYNSDDPLAECLVYVYIKRGQVRVGNVNDQTAELRLNGS	536
Db	791	GIMFQM-----DNHLPVNLNLEDPQSEMLLYIKIGETTVGKYKXPNSSHDILQSGV	843
Qy	537	KLKEHCTFENVNVVYVYENKAAVWVGVRIDKPTKRSYRIILGDFHIFRRNPEE	596
Db	844	LIADHCTIKNFGGVSYIIPVGEAKTYVNGKHILEITVLRHGDVYILGGDHFRRNHVE	903
Qy	597	ARARQROSLIRHSYNSQSGSPAPRGHRIRLSKASGDADGSRSDSPRPHRGKDSWF	656
Db	904	VQCKR-----PSGRDTPISSEGRPE-PAKNELMLAQRSSQLEAEIK	944
Qy	657	YAREEASAILGLDQKISHLTDELD---ALFDDVQKARAVRGVLEDNEDSDOSSPEV	713
Db	945	EALQKAEEMWQGIQIKEMAQOELSSQKAAVE--SKIALLELREBQGRKKMEI---	999
Qy	714	RDKWMSGTIDNSSLDTAITPGTSPASDDGALFEGDKSKQ-----DASVADVE	765
Db	1000	NNQKANHKEEL-----EKARQHLQOEIYVYKRRKLEMET	1033
Qy	766	LROQO-----AQMEALTKAKOE	783
Db	1034	LATQKALEHDSIRHARILEALETEKOK	1060
RESULT	15		
ID	009997	PRELIMINARY:	PRT: 928 AA.
AC	009997		
DT	01-NOV-1996	(TEMBLrel. 01, Created)	
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)	
DT	01-JUN-2001	(TEMBLrel. 17, last annotation update)	
DE	PUTATIVE KINESIN-LIKE PROTEIN R144.1 IN CHROMOSOME IIT.		
GN	R144.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
NC	STRAIN-BRISTOL N2;		

RA Favello T.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterton R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA All M.Y., Khan S.T., Siddiqui S.S.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U23515; AAK21455.1; -.  
DR EMBL; AB023497; BAA86882.1; -.  
DR HSSP; p17119; 3KAAR  
DR InterPro; IPR000253; FHA\_domain.  
DR InterPro; IPR001752; Kinesin.  
DR Pfam; PF00225; kinesin; 1.  
DR PRINTS; PRO0380; KINESINHEAVY.  
DR SMART; SM00240; FHA; 1.  
DR SMART; SM00129; KISC; 1.  
DR PROSITE; PSS0067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW ATP-binding, Coiled coil, Hypothetical protein, Microtubules  
KW Motor protein.  
SQ SEQUENCE 928 AA; 103589 MW; CA174A03AFCFEPA CRC64;

Query Match	28.5%	Score 1147	DB 5	Length 928
Best Local Similarity	38.8%	Pred. No. 2.9e-58		
Matches 281	Conservative 116	Indels 217	Indels 110	Gaps 19
Oy	1	MSGGNIKVVAVRPFNAREIDRGAKCIYRMGNGNTILTPPGAEKAKKSGKTIMDGP	60	
Db	1	MKGGSIIYAAVRFPNDREKTRCKLYEMDEETTYIRDKTDE-----K	48	
Oy	61	AFADRSYSED-----KNA-----PNARQEDLFQDGLVPLLDNAEFGYNCIFAYQ	109	
Db	49	RETYHSYWSHSGFSEKKNKGYLEPDPHADRDRAVEDRGSVLNNAMAGYNCISLFAYQ	108	
Oy	110	TGSGKSYMSMGKRGHGYIPRICOMFRRINLODKNITCIVESYLETIVNERPDLN	169	
Db	109	TGSGSYSTYVFGKNNKGIVPIYCEELFKIAD-NKKNNQFEYFYSMATEYEKADLLS	167	
Oy	170	--PSTKGNLKYREHDSPTYVEDLAKLVVRSFOETENLMDGCKNARKYATVATNMNETSRS	227	
Db	168	STPPKGGKGLYEHKPNFGYVLELTPVPSNEKTEAKIEEGTKSRITATQNMNTSSRA	227	
Oy	228	HAVFTLLTQKHDETKMDTEKAKISLYDLAGEBRSATSTGATGARLKEGAEINRSLST	287	
Db	228	HTIVATITNÖK--SSQOAGGTSMKKSEINLVYDLAGEBRSQSAATEDRKEGLVINOSSLT	286	
Oy	288	LGRTVLAALADSSGQKKNQVLPRYDSVLTWLLKISLGSNSKTAIAIASPIDINFEETL	347	
Db	287	LGRTVIALHDGSKAGSKKQTQIPYDSVLTCLLKNALGNSKTMITAIASPIDINFEETL	346	
Oy	348	STLRADSAKRRIKNAVYVNEPDNAPMIRELKEALQALNSKLOSGGGGAGCGSGPVBE	407	
Db	347	STLRADRAKRSIKTAAVYVNEPDNAPMIRELKEALQALNSKLOSGGGGAGCGSGPVBE	393	
Oy	408	SVPTDTLEKQVLSIQDPDATYKKRSKAEIYBOLNOSSELYLDNMTWMEKTLAKTEETHK	467	
Db	394	-----DASNEELEK--LRQQLLENQEMEMEKSMOOKIA--EBAK	431	
Oy	468	EREALLEEGLISIEKFGVGYRHSKEMPHLVNLSDDPLLAECFLVYNIKKPQTRVAGVNDPT	527	
Db	432	HASGASKEVEMEARK-----KKMCHLMNLEDNPALTINVIHFIPGYESVVGKPPSS	483	
Oy	528	QAEIPLNSKTLKECTEEN-VDNVYTIYP-NEKAAYVANGYRIRDKPRLBSGRIILIGD	585	
Db	484	GNEIYMGSLTIPQVHTLKDGNNGNIIHLSPCSEDDIDITNGRPVGEQTLQDNDRVFEGG	543	
Oy	586	FHIFRPNPEEARARQEOSLLRHSVTVNSQLSSPAPGRHDFRLSRAGSDADDSRSD---	642	
Db	544	NHLVYENPDKTKG-----TRTQITYENAAQALFAONNAAALISGRG--LGGSKRPDLITL	593	

Thu Jul 25 08:37:55 2002

us-09-235-416-1.rspt

Page 12

```

QY      643  -----SPLPHRGRKMSDWFYARREASATLGLDOKIS-HLTDELALFDYQKARAVR 696
Db      544  EEELMSTLP-----LVORANMATELGRNVKFEIVLVSPEM-----RGLTIS 654
QY      697  GLIVE 700
Db      635  GLTE 638

```

Search completed: July 25, 2002, 05:48:42  
Job time: 256 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:47:27 ; Search time 3312.9 Seconds

(without alignments)  
14856.837 Million cell updates/sec

Title: US-09-235-416-2

Perfect score: 2352  
1 atgcggcggtggaatat.....aaacagcgaacgaggaattc 2352

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_inu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	474.8	20.2	6972	9	HAAXTRSYV	X90840 H. sapiens m
2	463.8	19.7	7401	3	AF247761	AF247761 Drosophila
3	437	18.6	5421	10	MUSKIF	D29951 Mouse mRNA
4	421	17.9	5739	10	AB023656	AB023656 Mus muscu
5	419.4	17.8	4862	10	MUSKIFB	D17577 Mouse mRNA
6	398.4	16.9	5575	9	AY043362	AY043362 Homo sapi
7	398.4	16.9	6816	9	AF257176	AF257176 Homo sapi
8	382.2	16.2	4082	9	HS091329	U91329 Human kines
9	382.2	15.4	6217	9	AB014606	AB014606 Homo sapi
10	361.8	15.4	4117	10	RNAI696	AJ000696 Rattus no
11	361.6	15.4	3303	10	AB074017	AB074017 Mus muscu
12	352	15.0	5170	3	CELD104A	M58582 Caenorhabdi
13	352	15.0	7107	3	DM081788	U81788 Drosophila
14	346.6	14.7	6618	10	AB037923	AB037923 Mus muscu
15	328.2	14.0	5700	9	HSR291578	AJ291578 Homo sapi
16	328.2	14.0	5773	9	AY014404	AY014404 Homo sapi
17	328.2	14.0	5780	9	HSR291579	AJ291579 Homo sapi
18	328.2	14.0	5878	9	AY014403	AY014403 Homo sapi
19	318.8	13.6	9924	10	AF090190	AF090190 Mus muscu
20	317.2	13.5	5743	10	AF131865	AF131865 Mus muscu
21	316.6	13.5	8776	9	AB017133	AB017133 Homo sapi
22	316.2	13.4	10682	6	AX039604	AX039604 Sequence
23	314.6	13.4	8743	9	AF279865	AF279865 Homo sapi
24	311	13.2	6325	3	AF245277	AF245277 Dictyoste
25	307	13.1	561	8	CNS01C94	AL115232 Botrytis
26	300.6	12.8	2067	10	AF083331	AF083331 Rattus no
27	241.8	10.3	5532	3	AY070884	AY070884 Drosophila
28	235.4	10.0	2819	9	AB023497	AB023497 Caenorhab
29	215.6	9.2	130049	8	NC17E5	AL513467 Neurospor
30	193.4	8.2	3432	10	AB008667	AB008667 Mus muscu
31	192.6	8.2	2282	10	MOSKIF3	L12645 Mouse mRNA
32	192.4	8.2	2213	3	SUSPKIN2A	L16993 Strongyloce
33	189	8.0	6586	6	AX333633	AX333633 Sequence
34	189	8.0	6586	6	AX334308	AX334308 Sequence
35	189	8.0	6586	6	HUMORFW	D26361 Human mRNA
36	189	8.0	172148	3	LMFP696	AL512293 Leishmani
37	188	8.0	451	9	AF004425	AF004425 Homo sapi
38	187.2	8.0	4021	9	AB046810	AB046810 Homo sapi
39	184.2	7.8	2391	3	AF149285	AF149285 Caenorhab
40	178.2	7.6	69523	2	AC094844	AC094844 Rattus no
41	177.8	7.6	3319	3	LEIKINLIKE	L07879 Leishmania
42	177.8	7.6	3319	6	AR071970	AR071970 Sequence
43	177.8	7.6	3319	6	I11729	I11729 Sequence 2
44	177.8	7.6	3319	6	I18957	I18957 Sequence 1
45	176.8	7.5	453	5	MSU34655	U34655 Morone saxa

#### ALIGNMENTS

RESULT 1	HAAXTRSYV	6972 bp	mRNA	linear	PRI 28-MAY-1996
LOCUS	HAAXTRSYV				
DEFINITION	H.sapiens mRNA for axonal transporter of synaptic vesicles.				
ACCESSION	X90840				
VERSION	X90840.1	GI:1212916			
ATSV	gene; axonal transporter of vesicles.				
SOURCE	human				
ORGANISM	Homo sapiens				

REFERENCE	Eukariyota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	Furlong, R.A., Zhou, C.Y., Ferguson-Smith, M.A. and Affara, N.A.
TITLE	Characterization of a kinesin-related gene ATSV, within the tuberos sclerosis locus (TSC1) candidate region on chromosome 9Q34
JOURNAL	Genomics 33 (3), 421-429 (1996)
MEDLINE	96299637
REFERENCE	2 (bases 1 to 6972)
AUTHORS	Furlong, R.A.
TITLE	Direct Submision
JOURNAL	Submitted (16-AUG-1995) R.A. Furlong, University of Cambridge, Dept of Pathology, Tennis Court Road, Cambridge CB2 1QP, UK
COMMENT	Overlaps with M78444, M78705, T07754, T15633 and T77291.







[illegible]

```

VERSION      D29951.1  GI:976234
KEYWORDS     kinesin family protein KIF1a; monomeric motor; neuron-specific
              kinesin superfamily motor; axonal transport; organelle transport.
SOURCE       Mus musculus
ORGANISM     Mus musculus
REFERENCE    1
AUTHORS      Yamazaki, H.
TITLE        Direct Submission
JOURNAL      Submitted (15-APR-1994) Hiroto Yamazaki, University of Tokyo,
              Faculty of Medicine, Department of Anatomy and Cell Biology, 7-3-1,
              Hongo, Bunkyo-ku, Tokyo 113, Japan
              (E-mail:Yamazaki@kinesin.kaiabol.m.u-tokyo.ac.jp,
              Tel:3-3812-2111(ex.3336), Fax:3-5689-4856)
REFERENCE    2
AUTHORS      Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. and Hirokawa, N.
TITLE        The neuron-specific kinesin superfamily protein KIF1a is a unique
              monomeric motor for anterograde axonal transport of synaptic
              vesicle precursors
JOURNAL      Cell 81 (5), 769-780 (1995)
FEATURES     95292344
SOURCE       Location/Qualifiers
              1..5421
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /tissue_type="Brain"
              121..5208
              /note="121-2122: highly homologous to KIF1b; 331-1174:
              kinesin type motor domain; 4834-5163: PH domain"
              /codon_start=1
              /product="kinesin family protein Kif1a"
              /protein_id="BAA06221.1"
              /db_xref="GI:976235"
              /translation="MAGASVAVAVRPPNREKRSRSCIKIOWSGSTTTIVNRKQKRE
              ETPSPFSFDIYSHGTSHPEDININASQKQVITDISEEMQDAFESTINVCITFVGTGAGG
              KSYTMGKQEKDQGGIIPQCEIDLFSTRNDTIDNMSYSEVSEATYCEVRNDRLNP
              KNKNLRVREHPDLIPYVEDLSKLAVSYNDIDIDLMGSKNKPVTAAITNNKTSRSH
              AVFNIIIFQKRHDAETNTTTEKVKISLVIDLAGSERADSGAKGRLKEGANINKSLT
              TLGVNIFSLAEMDSQGNKKKKKTKDFIPYDSDVLTMLRENGNSRTMAVALLSPA
              DINDETSLTLYVADRAKQKQIINDEBPNQIIRELIKDVPRLRLIYAOGIGDITL
              DMNATVIGMSPSSSLSAISRAVSSTIHPILLARPSEAEIELKTEKTIANTLENT
              WEKLRTRERAIEMREALAMGYAMKREDGTLVFSRQKPHLVNINDELMSECLTL
              YIKDQTVRGREDIAERKQDVLVSGHFKTECHKVIRNEDSRSGAVVLTLECEADYTV
              VNGKRVETPSILRSNGNRIIMKGSVFRPNHEQARKORERPCAEPAEPVMAFAOR
              ELKQKQIDMKQKMEORLOELDOYRREBEATVILIQORLDLSKALQKQKDSYTY
              YPEINEEBEEDDEWOTRECELDALNAFKWMKYOTFSLDMLMGALFLKANAIST
              VELTKVOYQFVLLTDLTLYSLPDDLPPEAKKQRETRPPRTIVAYVEDODKQANH
              YMLTEKLRORLDIRREMYDRAAEVSSVEDCDVMVYGQEPFPRPMPFLVGAAPVY
              LSNLTLYPVIYHRAIYISKEKVEKGLFVANOAISADEEADPYSYSGROSTAKISD
              DQHEKQSESCPVAGMSRSGTSQEEELIYVGQOGADGASADVENNNKTSVAPRPG
              LMDSPKRAALDGLPDLTALDLRLGSLSTFTFRVTVQAQASISAEVADICQNFNIRHDE
              AFSTPEKLPKNGRGPPLGIFHYONAVTAVTSKFIYIKSOPITVEFVGHYOHPRLPC
              KDVLSPKTSRRHFRPWPVLSKVPATKLSMTPTSPGSPGCKKDLVYEICELNAC
              GDYIPAVUDHKGACMGTEFLHOGIQRITVTLIHETSGSHIMKKEVRLVYGRINPPE
              TDEALDNLILSLINLSSGYVHPADQDRVFNQNTPTTFPEELAMSSMINSILNV
              TPVREKITYMTLSATIEKENCQCPAVITIKDCQVYISDAKIPASRSTRNLFGSGSLA
              TEGKRVKELSLQEHVADGSPQMGRRRRVLDVTSVAYVGNENLGAEMPRSDSLTL
              DHQWELKSLSLQIEVKTRHYALLRELEKTRRRVGPVLLSPASSEDESSSSGASPS
              LSGASGSPSLPAPNERRELAVKCLRIIMHFNENEYHSHVCIASSEKLSDEKVTMLP
              RDPSPSLGATITPSTSCPLIEGRGADGVPPQCSRPASPEKLPDELQKTSK
              SPVATETERKEPQRLVLDIOEIRVSPIVSKGKYLHLEPTAGMARFVYVRPVAYV
              MYNKDQTVREVTNLSTAOVEYSEDDQAMIKTPTNAVCTEHNIGILLQANSKDMHD
              WLYAFNPLLACTISKLSRRSSAMRV"
BASE COUNT   1324 a 1593 c 1467 g 1037 t
ORIGIN

```

```

123 GCGCCGGGCGCTCTGTAAGGTGGCGGTGGTCCGCCCTTCAACTCCCGGGAATGAG 182
OY 66 ccgtgagcaaatgtatgtgtagatgaaagaaatcaaatcaatccatccctcc 125
DB 183 CGAGACATCCAAAGTGCATTCATTCAGATGCTGTGAAAGCACACACTATTGTAAACC 239
OY 126 ggtgtcgaaagaaagcggttaaaagtgtgcaaaactataltgtgtgcccgaagcalt 185
DB 240 -----CAACAGCCCAAGAGACACCCAAAGGCTT 269
OY 186 tgcgttcgacgttcgtatgttgcctttgac---aagaatgtcccaactatgagagaa 242
DB 270 CAGCTTTATTACTCTCTACTGTGCATACCTCGCCGGGAGCATCAACTATGACCTCTCA 329
OY 243 gaaagacctatccaaagctcgcgaagtcgcgtctcgtgtgatagtcaatcaaggtataa 302
DB 330 CAAGCAGGCTCTACAGGACATTGGGGAAGAGATGCTACAGCATGCCCTTGAAGGCTACAA 389
OY 303 caattgtatcttcgctacggtcagaccggttcggtgcaagctcattcaatgtatgtgcta 362
DB 390 TGTGTGATCTTGTGCTATGGGACAGCGGGGCGCAAGTCTTACCATGATGGGGA 449
OY 363 tggcaag-----gagcatgtgcgtatcccgcgatttgcagagacatgttccggtat 416
DB 450 ACAGAGAGAAGACACAGAGGATTCACAGCTTCGGAAGACCTCTCTCGAAT 509
OY 417 taatgaactgcagaaagaaagaaactcaactgtcacccgtcgaagtgttctgatttgaat 476
DB 510 CAACGACACACCAACGAC---ACATGTCTACTCTGTGGAGGTGAGCTACATATGAGAT 566
OY 477 ttacaatgaagagtgagagacttgtcaatcgtcgcagaaagggaaatcctcaagttccg 536
DB 567 TTACTGTGAGGAGGACCGGACCTCTGTGAACCCCAAGAAAGGAACCTCGTGTGAG 626
OY 537 agaaacccgttcgaccggtccctacgtgtgaggaactgtgcgaagtgtgtcgtatcalt 596
DB 627 GGAGACACCTCTGCTGGGGCTTATGTGTGAGGACCTCTCAAGCTGTGGCTAACTCCCTTA 686
OY 597 ccaagaatcgaataatcatalgagtagaggaagaaagcaagaaaggttgcgcgcacaa 656
DB 687 CAATACATCCAGGACCTCTATGACTCAAGGAAACAGCCAGACTGTGGCCGACCAAA 746
OY 657 catgaacgagacatccagtcgaatcccaagcggtcttcaacttgaacttgcgaaagtgt 716
DB 747 CATGATGATGACCAAGCAGCGCTTTCACGACGATTCATCATCTTATACCAAGAAAGG 806
OY 717 gcatgataagagagacaaatgtgacacagagaagtgtcgaagaatcgtctgttagatt 776
DB 807 CCATGATGACAGACCAACATCACACACAGAAAGTGAACCAAAATCAGCCTGTGTGACCT 866
OY 777 ggcggttcgtgagcgcaagtcgacccgagctactgagcgacgtcaatgaagaggtgc 836
DB 867 GGCAGGCAAGTGAACGACGACTCCACAGAGACCAAGGCAAGCAGCACTAAAGAGGGGCG 926
OY 837 agaatcaacacgtcacttcgacccctaagtcgtgtgaltgtcagcgctatgcgatatgtc 896
DB 927 CAACATCAACAAGTCTTTGACCAACCTGGGAAGGTCAATCTGCTGTGGTGAATGGA 986
OY 897 gtccgga-----aaacagaagaagaatcagttagtaacttaccagagattc 941
DB 987 CTTCTGACCCCAACAAGAACAGAAAAAGAAAGACAGATTTTCATCCCTTACCGAGACATC 1046
OY 942 ggtactgagtcgttcgtgaagactccttgggaagcaactcgtatgaagcagatgttc 1001
DB 1047 CGTATTGACATGGCTTTCGCCGGAACCTCGTGTGCACTCAAGACAGCCATGTGGC 1106
OY 1002 cgcacattcgcgtgtgataacttgaagaagactcagtaaccttcagataltgcgga 1061
DB 1107 ACCCTTGAGACCTCGGAGACATCACTATGATGAGACCTTCACACACTGAGATATGCAGA 1166
OY 1062 cctctgaaaggaatcaagaacacggtgtgtcaatgaagcccgaaagcgcgatgtgt 1121

```

```

DB 1167 TCGGCCCAAAACAAATCCGCTGCATGCAATCATCAATGAAGACCCCAACAGCTGAT 1226
OY 1122 ccgcagttgaaaggaagaaactcgcgcagcttgaaggaagcaactccagacagtggtgag 1181
DB 1227 TAGAGAACTGMAAGATGAGTGAACGAGCTGAGAGACCTGCTGTATGCCCAAGTGTGG 1286
OY 1182 tggaggtgtgcaagaggttctgtgcggccagtgagaaactgtaccgcccagacagcc 1241
DB 1287 CGATATCACTGACATGACCAATGCCCTGGTGGGACAGACCCCTCATCTTCACTCAAC 1346
OY 1242 gctcgaagaagaatcgtgtcgtatcagcagccggtgtgcagatgaag---aaatgag 1298
DB 1347 CCGTTCACGCCCGTGCAGCCTCTGTGTCCAGTCTCCATGAGCGCATTTTGTCTCCAG 1406
OY 1299 caagcagaatcgtgtgagcaacttgaaccaaagtgagaaactcatalcggatcatalca 1358
DB 1407 CAGCGAGAGGCTATTGAAAGCTGAAAGAAACAGAAAGATATTCGTCAACTCAATGA 1466
OY 1359 gacctggaaagaagaactgtgcgaagccgaagaaatccacaagaaagaaagcggt 1418
DB 1467 GACCTGGAGAGAAAGCTGCGGCGAGACAGAAAGCATCCGATGAGAGGAAAGCCTTGGT 1526
OY 1419 cgaagagctgggt-----atcagcatcgaaagaggttgtgtgaccttaccatccaa 1472
DB 1527 GCGTGAGATGGGTGTGGCCATGAGGAGAGATGCTGTACCTGGGTGTGTCTTCCCA 1586
OY 1473 agaaatgacacatcgtatcaacttgaagcgatgactccttctgtcgtgtgtgttctta 1532
DB 1587 AAGACTCTCTCATCTTGTCAACTGTAATGAGACCTTTGATGTCCGAATGCTCTCTTA 1646
OY 1533 caaatcaagcccggtgcaagaaggtgtgaaacgtcaaccaaagaaatcaagcgaagaa 1592
DB 1647 CTACATCAAAAGATGGGTGACCAAGATGGGCAAGAAAGATGCAAGAGGGGCAACAT 1706
OY 1593 tgcctgaacggttcgaatcctgaagaacacgtacgtacgttgaagaatgtg----- 1644
DB 1707 TGTCTGAGTGTGACCTTCAAAAGAGAGACCTCATCTTCCGAGTACCTCCGAG 1766
OY 1645 -----gacaacgtgtgtacacatcgtgcgaacgaagaggtcgtcgtatgttgaag 1697
DB 1767 AGCGGAGAAAGCTGTGTGACCTGTGAGCCCTGTGAAGAGCAACATATTCACAGG 1826
OY 1698 cgtgcgaatcgaaacactactcgtccgcgagcggtcagatcatccgtgtgagatt 1757
DB 1827 CAAGAAGATCACTGACGCCAGCATCTTACGGTCAAGAAACCCGATCATGTGGAAAG 1886
OY 1758 ccaacatttgcattcaacacatccggaagaagctgtgtgcagcg 1802
DB 1887 CCAGTGTTCGGTTTAACCAACCTGAGCAGGCCGTCAAGAGCG 1931

```

```

RESULT 4
AB023656 5739 bp mRNA linear ROD 24-FEB-1999
LOCUS AB023656
DEFINITION Mus musculus mRNA for KIF1B-beta, complete cds.
ACCESSION AB023656.1 GI:4512329
VERSION AB023656.1
KEYWORDS KIF1B-beta.
SOURCE Mus musculus (strain:ICR mouse) 4 weeks CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE Nakagawa,T. and Hirokawa,N.
AUTHORS Identification and characterization of a new kinesin superfamily
TITLE KIF1B-beta
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 5739)
AUTHORS Nakagawa,T. and Hirokawa,N.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-1999) Terunaga Nakagawa, Grad. School of
Medicine, Univ. of Tokyo, Department of Cell Biology and Anatomy;
7-3-1 Hongo Bunkyo, Tokyo 113-0033, Japan

```

(E-mail: hirokawa@m.u-tokyo.ac.jp, Tel: 81-3-3812-2111 (ex.3336),  
Fax: 81-3-5689-4856)

## FEATURES

## source

Location/Qualifiers

1. 5739

/organism="Mus musculus"

/strain="ICR mouse"

/db\_xref="taxon:10090"

/dev\_stage="4 weeks"

224..5536

/gene="Kif1b"

224..5536

/gene="Kif1b"

/note="kinesin superfamily"

/codon\_start=1

/product="Kif1b-beta"

/protein\_id="BA05243.1"

/db\_xref="GI:4512330"

/translation="MSGASVKAVRVRPNRSRTSKESKCIOMQNSTSIINPNKPK

EAPKSPFYSYMSHTSPDPCFASQNRVYNDIGKEMLHAEGYVNCIFAYGQTGAG

KATYWMKQEEQAGIIPOLCEELPEKINDCNEMSESVGEVMEYICGRVRLINP

KKGKINRVREHPLIGPYEDISKIATVSTDIADMDAGNKAATVATMMNETSRSH

AAFTVTOEKODPEPNLSTEKYSKISLYDLASEADSGAGTGLAKNKISLT

TIGKVISALAEVSKKKKTDFTIPYRDVLTWLEKREMGSRITAMVAALSPADINDE

TLSTLYADRAKQIKCNAYINEDPNKLVLELKEVTRKLDLRAOGLDIDTSMGS

LTSPPSSCSLNSQVLTYSVTSIOERIMSPGGEALERLKESEKIIIAELNEMWEKLR

KTEATIMEREALIAEMGVAIREDGTLGVSPKPTPLHLVNLNDPISBCLLYIKDG

ITRFGADARERODIYLSGAIHKEECLFESRSNTEVLTLEPCERSYVNGKRV

AHPVQIRSGRITIMGNHVRFNHPQOVARERKTPSAEYSPVPEVTOVERSLAET

GIDMKOEMERLODEMLYKKEKEADLLEQORLYESLQALOVNRETSIAAET

EEDDEBEVPTOHEFELADOMARKMSHOFIPRTVAVEQDLKNGTPIWLSLK

VQRFVLDLTLSPVPELTPSEMEKTHDRFPRTVAVEQDLKNGTPIWLSLK

LKORLIDREMYDRAGVAVASADDSSTMTGSDPPTVDRHMKLVGSPITFGCVNE

RLADRTSSPFTASDITELADQODAMEDPDEAVDGTGADAGESELSIDGH

DPIYDRSPWFLVGRAFVYLSNLLYPVPLIAHYAISEKEVGFLEAVQALAADE

APYGSISIRSGTAKISFDEYFNOSFSAATRSGLSEELRYVGOSSEVISP

PEEYNRANDLISGTLIDCKMWMESSEPTGKIKGSAFTRVTVLOAGSSVISP

DIFQENFLRHDAESTEPKNNRGSPGLFHVONIAVEYTESPYDTIKRTPIYFE

VFQHYOQPHLQODINSPPOPSRRFPMPPLSKPVPAKTNMTKNTLIGOSMAY

DLVWEISETLEGEXIPAVVDHTAGLPCOGTFLHQGIQRIYVTHIHEKSELHM

KQRELVVGRIRNKPPEYDEAVASLANTISAKLSAAHSSTREFRFAVWDSLSH

NSLLINVTPIYGEKTYMTLSAYLEIDHCIOPAVITKQVAFYSRAKISPPSLNL

FGSGYSKSPDSNRPTGYIELSLCKMAPDGPOMGRKRYKIDNSVAREENAGRR

PRGDSLTLEHOMELKLELHEVEKTRHFLLRDLBRDSDPKSLSDLSVLSGGLS

TSTISQISETTFEESATITPSESGYDSAVESLVDEKELATCLOLHTHERES

QVHGSISDCKLSDISPIGRDPSVSSFSSTLTPSCTPSVDSNSESMDKTYLHPNR

AASPCOEFOIIVPTVEPYLAAGKNEELNLPDIEEVRAGSVSSKKYELHKEPL

SSNMAKHFVVRPRYFTIYNSKDPVERGIINLSTAOVESEDOAVKVPNTFAVCT

KRGVLLQALNGMDNMLYAFNPLLAGTIRSKLSRRCPSQPKY"

BASE COUNT 1490 a 1414 c 1501 g 1334 t

## ORIGIN

Query Match 17.9% Score 421; DB 10; Length 5739;

Best Local Similarity 56.8%; Pred. No. 1.4e-108;

Matches 954; Conservative 0; Mismatches 680; Indels 45; Gaps 8;

QY 166 atgagtgccgcaaggacattgctgcgtcgtatcttgcttcgcaag---aat 222  
DB 353 AAGGAAGACCAAAAGTCTTCACTTGTGACTCTACTGCTCCACACACTCGCCTGAA 412  
QY 223 gctcccaactatgagagacaagaacctatccaaagatctcgggtgccggtttgat 282  
DB 413 GATCCCTGTTTGCATCGAGATCGTGTCATGACATTTGAGAAAGAAATGCTCTTA 472  
QY 283 aatgattcaagagttataatacaattgtatcttcgctacgctcaagccggttcgggcaag 342  
DB 473 CATGCCCTTTGAAGGTTTATATGTCGTATCTTTGCCCTATGGGCAAGACTGCTGGGAAA 532  
QY 343 tccatttcaatgatgggctatgagcaagagag-----catggcggtgatacccgcgattgc 396  
DB 533 TCTCTACACGATATGTGGGAAACAGAAAGAGAGCAAGTGTGAATATCCACAGTGTGT 592  
QY 397 cagagatagtctccggttataatgaactgcaggaagacaagaactcacttcgacgctc 456

DB 593 GAAGAACTCTTTGAGAAAGATCAATGACAAATGTGAATGAAGAAA---TGTCTTACTCTGTAT 649  
QY 457 gaagttctgacttgtaaatattacaatgaagagatgcgagacttgctgatccgtcagaca 516  
DB 650 GAGGTAGGCTTACCAATGAAATTTACTGTGAGAGATGACGAGATTTTACTGAATCCCAAAAAC 709  
QY 517 aaggggaatctcaaggtccgagacacccgctgcagccgcccactcgttgaggacttgagc 576  
DB 710 AAGGTTAACTGGGTGTGGGTGAACACCCGCTCTTGGACCCGATATGTGAGAGATCTGTCC 769  
QY 577 aagctggtcgtcgatcatcattccaagaatccgaatcctaatctgatagtgagggccaagacc 636  
DB 770 AACCTGGCGGTCACTCTCTACCTGACATCGCTGACCTGATGATGATCTGGGAAACAAAGCC 829  
QY 637 agaaagcttgccgcaaaacatgaacgagacatccagctgacccagccgctcact 696  
DB 830 AGGACTGTGCACCTACCAACATGAATGACACAGACGCCGTTCTACCGCTGTGTTCCACC 889  
QY 697 ttgacctgagcgaagaagtgcatagtatgaagagaccaaataatggacacagaagaagttgcg 756  
DB 890 ATTGCTTTACCCAGAAAGAAAGAGATCCTGAGACGAACTTCTACTGAGAGGTACAGT 949  
QY 757 aagatcagctcgttagatcttgccggtctcgaagcgagcaacgctccacccgagactatgaa 816  
DB 950 AAATATGCTTGGTGTGATTCAGCAGGAAGTGAACGGGCTGATTCACTGAGTGCCTGCAAGGA 1009  
QY 817 gcgcgctgagagagaggtgcagagatcaacgctcacttcgacccctaggctgagatt 876  
DB 1010 ACTGCAATTAAGAAAGAGTGCAGAAATATCAATAGCCCTTCAACTTTGGGCAAAAGTCATT 1069  
QY 877 gcagcgctagcgatattgctgcgggaaaaacgagaagaatacgaatgaacttaccca 936  
DB 1070 TCGGCTCTGGGAGA--GGTAGATTAAGAAAGAAAGAACTGATTTATTTCTTATAGG 1126  
QY 937 gattcgttactgagctgcttcgaaagactcctctggagagcaactcgatagacccgcat 996  
DB 1127 GATTCCTGTACTTACTTGTGCTCTTCGAGAAATCTCGGTGGCAACTCCGCAACTGCATG 1186  
QY 997 attcgcgcatttcgctcgtcatattacttgaagaagactcagttacccttcgat 1056  
DB 1187 GTTGCTGCTGTCGAGCCGACCGCAGCATCACTACGATGAATCTTGACACACTGATAC 1246  
QY 1057 gcggaactcgcgaagcgatcaagaacacgacgctgctcaatgaagacccggaagcgcg 1116  
DB 1247 GCAATCTGTGAAAACAAATTAATGCAATGCTTTATCATATGAGACCCGAAATGCCAAG 1306  
QY 1117 atgattccgcgagttgaagaggaactcgcgagctggaaggaact-----ccagagc 1170  
DB 1307 CTGGTTCCGGAGTTGAGAGAGAACTGACCCGACTGGAAGACCTTCTGCTCGTCAAGGA 1366  
QY 1171 agtgtagagtgagagtgctgcagagagtgctgcggcgccagtggaagatcgtaccg 1230  
DB 1367 CTGGAGATATTTATTTGATACATCATGCGGCTCCCTCATCTTCATCTTCATGCTCA 1426  
QY 1231 ccgcgacgacgctcgcgaagaacaaatcgtgcgatttgaagcgcgagcgatgagcaagtc 1280  
DB 1427 CTTAATATCTAGGTTGCTTAACTCTGTGACAGATATTA---GGAGAGATATATGCT 1483  
QY 1291 aaaaatgagcagagcaaatcgttgagcaactggaaccagagatggaagctctatcggat 1350  
DB 1484 ACACCTGGAGGGAGGAGAGCACTTGAACGTCTGAAAGAGTGTCAGAAATCATATTGCTGCT 1543  
QY 1351 ctcaatcgaactcgggaagagaagcttcgccaagacgaggaatccacaagaagacgaa 1410  
DB 1544 TTGAATGAACCTCGGAGAGAGAACTTGTAAACAGAGGCGCATCAAGATGAGAGAGAA 1603  
QY 1411 gcggcgctcgcagagactggtatcagcat-----cgaagaagggcttgcgcccctac 1464  
DB 1604 GCCCTTGTGCTGAATGGAGACTTCCATACGGGAAGATGGGGGACACTTGGAGATTTC 1663  
QY 1465 cactccaagaagaatcccatctagctcaactgagcagatgactcctcttcgtgagtgat 1524



OY	937	gatttcgactacgacgtggtctcttgaaagacctccctggagagaaactcgatgaccgcaat	996
Db	1136	gatttcgactacgacgtggtctcttgaaagaaatcttgggggaactcccgaaatcgaaatg	1255
OY	997	attgcgcgcatttcgcctcgtcgatataactttgaagagacctcagtaaccctcgatat	1056
Db	1256	gtttgcgtcctcagaccacgcacatcactgcagatgaaacttttgagcactcgagatc	1315
OY	1057	ggagacctcgggaagcgaaatcaagaaccacgaagtggttaatgaagaccggaacgcgcg	1116
Db	1316	gcagatcgcgcaaaacaaatgaaatgcaatgcgttatcgaatgagaccgcgaatgccaa	1375
OY	1117	atgataccggaattgaaggaaggaactcgcgcacctggaaggaacaaact-----ccagac	1170
Db	1376	ctggttcgggaattgaaagagaaagaaacccagctgaagacactcttctcgtctcagga	1435
OY	1171	agtgttggaagtggaagtggtgtgcaggaaggtcttcgcgcgccagttgaagaatcgtacc	1230
Db	1436	ctggagagatattattgatacattcattgggggtccctcactcttccatcttcattgcctca	1495
OY	1231	ccgcgaacgcgcctcgaagaagcaaatcgtctcgattcagcagccgatatgcagcagtc	1290
Db	1496	ctttaattagtcagagtggtggccttaacatctctgaccacgatattca---ggagaggaatc	1552
OY	1291	aaatgaagcaaggcagaanaatcgttggaagcaacttgaaaccagaatgaaagctcatcg	1350
Db	1553	acacctggaggggagaaacccattgaaacgctggaagagatcagacagatcattcgctgac	1612
OY	1351	ctcaactcagaaccttggaagaagaagctgscgaagacgcggaagaatcacagaagaacga	1410
Db	1613	ttgaattgaaaccttgggaamaagaaacttctgtaaaacagagccatcgaattgaaagaa	1672
OY	1411	gcgcgcctcgaaggaactggtgtaacgaat-----cgaagaaggtctgttgccctaac	1464
Db	1673	gccttgttggtgcgaatggagatggatggcaatcaggaagatcggggggacattgagagtttc	1732
OY	1465	cactcagaagaatgccaatctagttaacttgagcagatgataccctctcttgctgaagtt	1524
Db	1733	tcaccttaaaaaacccacacatttctgtaaacctcattgaagaccattgaatgctggagtt	1792
OY	1525	cttgctcaacaacaaagccgcggcagaagaaggtgtggaagcgtcaacaagaataacaa	1584
Db	1793	ttgcctcattattacatcaagatggaatttacaaggttggccagacgaatcttgagcgcc	1852
OY	1585	gcgcgaatctcgtcgaacggttcgaagatcctgaaagaacactgtacgtttgaaaatgtg	1644
Db	1853	caggacatcagtcagtcagtcggccccaacattaaagagagacatttgccttttggagtgag	1912
OY	1645	gacacag-----ttgtaaccatcgtgcgaagaagaaagctgtctcatg	1689
Db	1913	aggagacacacgtcggaagattttgttgacctttggagccctgtgagcgcctcagaaacatc	1972
OY	1690	gtgaacgcgcgtcgaatcgaacagactactcgcctcgcgcgaagcgtctacagatacactg	1749
Db	1973	gtgaatggcamaagaggtggctcatcctctgttcagctgcctcagggaaacggtatcattc	2032
OY	1750	ggcgcattccacatttttgattcaacatcccgaaagaagcttcgtgcgaaacgcgcaaga	1808
Db	2033	ggttaaaacacattgttttctgttttaaacaccttgaaacagcccgagacaaagggaaaa	2091

RESULT	6					
AY043362						
LOCUS	AY043362	5575 bp	mRNA	linear	PRI 18-AUG-2001	
DEFINITION	Homo sapiens kinesin superfamily protein 1B (KIF1B)				mRNA, complete cds.	
ACCESSION	AY043362					
VERSION	AY043362.1	GI:15212239				
KEYWORDS						
SOURCE	human,					
ORGANISM	Homo sapiens					
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					

REFERENCE	1 (bases 1 to 5575)
AUTHORS	Park, M., Shin, H., Lee, Y. M., Moon, E., Choi, W. and Kim, W.
TITLE	Identification of the human ortholog of mouse Kif1b, a kinesin superfamily motor protein
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 5575)
AUTHORS	Park, M., Shin, H., Lee, Y. M., Moon, E., Choi, W. and Kim, W.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Biological Science, Ajou University, San 5 Wonchon-dong, Paldal-Cu, Suwon, Kyeonggi-do 442-749, Korea
FEATURES	Location/Qualifiers
SOURCE	1..5575
gene	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="1p36.22" 1..5575
CDS	/gene="KIF1B" 5..3466 /gene="KIF1B" /note="similar to mouse KIF1B; similar to KIAA0591 and KIAA1448 sequences found in GenBank Accession Numbers AB011163 and AB040881, respectively" /codon_start=1 /product="Kinesin superfamily protein 1b" /protein_id="AAK85155.1" /db_xref="GI:15212240" /translation="MSGSAVYAVRVPFNSRETSKESKCIIOGQNSTSIINPNPKP EAPSPFEDYSYMSHTSPEDPCFASORVYNDIGKEMLLAFEGYANVCIFAYGOTGAG KSYLMGKQESQNGIIPOLCELFEPNNHNDGNEMSYSEVSMETICEYRLLNLP KNKMLVRDHPLEPGTYVEDSLAVTSYSLVDIADLMQGNKARTVAATNMETISRS VYVYIVTFQKKHNETLATVYKSTSLVDIADGDRASFGAKTRLEGAIIKRSLLT TLGVSIALAEVSKKKRTDFIPYRDSVLFMLRENTNGSRTAMVALSPADINDYDE TLSTLRADRAKQIKCAVATIDEDNAKIVELKEVYRLKDLTAOGIGDIIDTSMGS LTSPPSSCLSSOYGLTSTVTSIOERIMSTPGGEALIRKESKIIIEINLTWEKIR KTEAIRMERRELLAEMGVATIREDDGTLGVSPKTPRIYVNLNEPLMSCELLIYKKG ITRGQDAERKQDIVLSGAHIKEHCITFNSERSNSEVLTLEPCRSERYVYGVKRV SOPVLRSNGRIITMGKNHVFENHPEDQARERKTPPEATPSEPVDTFQORELLEKQ GIDMKQEMERKILQEMELIYKKEEADLLQOGLDADSDGSDSRCESEKILITL SLRKLEPPSKLOTITVKGCGSPGSKRKREPIKMYIIPORRLSKDSKWTISDLKIOAV KEICEYVALINDFRHSROETELATVKKKEIKAGKDPNDRDSRAVARADVPMTGCV GDEITDEVYATGKSTVDYDLKTHIDLDELIDLEPKKQNNKKEDEIYLRKMLKMK VLPLIGSQEQSPGSHAKPEVGAVSSTSENNYSKQDNCELAEEVVSQLMNDDPAF RGRRLRMWRQEQIRFKNIQOQETIKQLRQVNPRTIPPEKRRKRPFFKSPKRNISM SPGHKIIITEDEVLEIRLIPKDEARKGNKESSQGGKAFKDPQPMGSGKSDPH IOVKHIIINNOQEPOLRMRNSNSLNNOGPKSTRQASASAEISLNSHGHPADVQTFQ AKRIIHOHROSICYNNTGCGLEGMAISYQKQDKPKSHCSQFVTPPMRRQFSAPNLK AGRETTY" 26..1072 /gene="KIF1B" /note="Region: KISc ATPase domain; catalytic domain of motor" 35..1120 /gene="KIF1B" /note="Region: motor domain" 1502..1750 /gene="KIF1B" /note="Region: FHA domain" BASE COUNT 1720 a 1101 c 1283 g 1471 t ORIGIN
Query Match	16.9%; Score 398.4; DB 9; Length 5575;
Best Local Similarity	55.9%; Pred. No. 4e-102;
Matches 941; Conservative	0; Mismatches 656; Indels 45; Gaps 8;
157	aaactatgatgatgagccgaaggcatlltgcitgcatacgcgtatgtgcttgcac 216
125	AAGATCCAAAGAAAGCAAGCTCCAAAGCTTCGACCTTCGACATFTCTACTGGCTCATACC 184
217	---aagaaatgcccaactatgcagagacaggaagaactatccaaagatctcggagtcgccg 273
185	TCACCCGGAAGATCCGTTGTTTGGACATCCAGAAACCGCTGTGTCAATAGACATGGCAAGAA 244

274 ctctcgtgaataatgcaatgaaggtatatacaaatgtatcttcgtctacgtgcaacccgt 333  
245 ATGCTCTTACAGCGCTTTGAGGGATATATGTCGTATTTTGGCTATGGCAGACTGCT 304  
334 tcggcgaagctcattcaatgatggtctatgcaagg-----agcatgctgtatccg 387  
305 GCTGGAATAATCTTATACATGATGGTAAACAAGAAAGAAAGCAGCTGGCATCATCCCA 364  
388 cggatttcgcaagacatgtccggtctatgaactcagaagaagacaagacctact 447  
365 CAGTTATGTGAAGACTTTTGTGACAAATCATGACAACTGTATGAAGAAA---TGTC 421  
448 tgcacgcgtcgaagttcgtctacttgaatattcaatgaacgaatgctgagacgtcgtgaat 507  
422 TACTCTGTAGAGGTGAGTACATGGAATAATTACTGTGAAGAGTACGAGATTGTGTAAT 481  
508 ccgtgcacaagaagggaatctcaaggtccgagaacacccgtcgacccggccctcagtggag 567  
482 CCMAAACAAGGTAATTTGGCTGTGGTATCACCCCTTCGTGTGACCCCTATGTGGAG 541  
568 gacttggcgaagctgtcgtcgtcatctcaagaatcgaagaatcgaatctatgatggagc 627  
542 GATCTTTCCAGTGTGGCAGTTACTTCTTACACAGACATTGCTGACCTCATGATGCTGGG 601  
638 aacaaagccgaacggtttgcccgaacaacatgaacgaagacatccagtcgacccagcgc 687  
602 AACAAAGCCAGACAGTGGCAGCTACAAACATGAAATGAACAAGTGGCCGTGCCAGCTT 661  
688 gtctcacttgaacttgaacgaagaatgagcatgatgaagagacaaatgaacacagag 747  
662 GGTATAGAGATTGTCTTACCACCAAGAAACAGATGAATGACCACTTCCTTCCACTGTG 721  
748 aaggttcgcaagatcagtcgtgtagatttgcggttctcgaagcgaacgtccaccgga 807  
722 AAGGTTAGTAAACACAGCTGTGGTGTGATCTACGAGAAAGTACAGCAGCTTCAACTGCT 781  
808 gctactggagcgcgaactgaagaagggtgcagaagatcaaacgcgtcacttcgaacctagt 867  
782 GCCAAAGGAGACTCGCTTAAAGAGAGAGCAATCATCAATATCTCTTACAACTTTGGGG 841  
868 cgtgtagctgagcgcgtcagcagatatgtctgcggaacagaagaagatcagtagta 927  
842 AAAGTCATTTTACGCTTGGCCGA---GCTGAGTAAAAAAGAAAGAAACAGATTTTAT 898  
928 ccttacgcgaatctgtagcagcgttgcgttcgaaggaactccttcggaggaacactcagat 987  
899 CCTCTACAGGAGATTCTGTACTTACTTGGCTCTTCGAGAAATTTAGTGGCAATTTCCGG 958  
988 accgcacatgattgcgcgaacttcgcctcgtatataactttgaagagaactcctcagtaac 1047  
959 ACTGCAATGCTGTGCTGTGAGCCCGCGGATATCATCACTAGATGAGACTTTGAGACACT 1018  
1048 ctctgatatgaggaacttgcgaagcgaatcaagaacacgcagtggtcgaatgaagcccg 1107  
1019 CTGAAATATGCAAGATGCTGCAAAACAAATTAATGCAATGCTGTATTCATGAGGCCCC 1078  
1108 aacgcgcggaatgatcgcgaagtgaagaaggaactcgcgcgaactgagaggaacaaac---- 1163  
1079 AATGCAAAAGTGTGCTGTAATTAAGAGAGAGTACACAGGCTGAAGAACCTTCTTCGT 1138  
1164 ---cgaagcagtggtgagagtgagagtggtgcagaagagttcgtgcggcgaatgagagaa 1221  
1139 GCTCAGGGGCTGGAGATATTTTGTATCATCATCATGAGGCTCCCTCACTTCAATCCCATCT 1198  
1222 tggtagccgcgcgcgaacgcgcgtcgaagaagcaaatcgtgtgattcagagccgagtgcg 1281  
1199 TCTGTCTCACTAGTAGTAGGTGGCTTGACGCTCTGTACCAAGTATTTCA---AGAGAG 1255  
1282 acagtcgaagaataatgagcaagcagaatcgtgaggaacactgaaccagagtgagaagctc 1341  
1256 ATCATGTCTACACCTGGAGAGAGAGAAAGCTATTGAACGTTTAAAGAAATCAGAGAAAGATC 1315

QY 1342 tatcgggattcctaatacgaactcgggaagaagacttgcgcaagaacgcggaagaatccacaag 1401  
DB 1316 ATTTGCTGAGATTGAAATGAAATCTTGGAAAGAAAGAACTTGTATTAACAAGACCAACAAGT 1375  
QY 1402 gaacgagaagcggcgtcgtcgaagagcgtggtatcagat-----cgaagaagcttgtc 1455  
DB 1376 GAGAGAGAGGCTTTGTTGGCTGTGAGATGGAGTGTGCAATTCGGAAAGATGAGGAACCTTA 1435  
QY 1456 ggccttaccaactccaaagaagaatcagacatctagtaacttgaatgagatgactcttcgt 1515  
DB 1436 GGGGTTTCTCACTTAAAGAACCCACATCTTGTATTAACCTCATATGAAGCCACATTAATG 1495  
QY 1516 gctgagtgcttgcgtcacaacatcaagcccggcggaagaaggttgcgaacgtcaaccaa 1575  
DB 1496 TCTGAGTGGCTACTTATTATTAATCAAAAGATGATTAACAAGGTTGGCCAAAGCATGTGT 1555  
QY 1576 gatacaagaagcgaagaatcgtcgaagcgttcgaagaatcctcgaagaagaacgttgaagtt 1635  
DB 1556 GAGCGCGCCGAGACATAGTGTGAGCGGGGCTCACATTTAAAGAAAGCATTTGTATCTTC 1615  
QY 1636 gaaatgtgacaacg-----tctgacacatcgtgcgaacgaagaagct 1680  
DB 1616 CGAGTGTAGAGAAAGCAACAGCGGGGAAGTTATGTGATCTTAAGCCCTGTAGCCGTCA 1675  
QY 1681 gctgtcaatgtgaaacggtgtgcgaatcgaacgaactcgtcgtcgcgagctacag 1740  
DB 1676 GAAACCTAGCTAATTAATGCAAGAGAGGTGTCCCACTGTTCAGTGCCTCAGGAAACCGT 1735  
QY 1741 atcattcgtggcgattccacatttctgattcaaccatccggaagaagctcgtgagaa 1800  
DB 1736 ATCATCATGGGTAAACCAATGTTTCCGCTTAAACCAACCGGAAACGACGAGCTGAG 1795  
QY 1801 cg 1802  
DB 1796 CG 1797

RESULT 7  
AF257176 6816 bp mRNA linear PRI 02-MAY-2001  
LOCUS AF257176 Homo sapiens kinesin superfamily protein KIF1B mRNA, complete cds.  
DEFINITION AF257176  
ACCESSION AF257176.1 GI:13925306  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 6816)  
AUTHORS Yang,H.W., Takita,J., Chen,Y.Z., Soeda,E., Piao,H.Y., Hashizume,K.  
and Hayashi,Y.  
TITLE Cloning of human kif1b gene that maps at 1p36, which is  
homozygously deleted in neuroblastoma cell line NB1 and mutated in  
some other neuroblastoma cell lines  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6816)  
AUTHORS Yang,H.W., Chen,Y.Z., Takita,J., Soeda,E., Piao,H.Y. and Hayashi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-APR-2000) Pediatrics, The University of Tokyo, Hongo  
7-3-1, Bunkyo-Ku, Tokyo 113-8655, Japan  
FEATURES  
source location/Qualifiers  
1..6816  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.2"  
154..5466  
/note="Kinesin related protein"  
/codon\_start=1  
/product="Kinesin superfamily protein KIF1B"  
/protein\_id="AAK49332.1"  
/db\_xref="GI:13925307"  
/translation="MSGASVKAIVRRPNSRSTKESKCIITIMOGNSITINPKPK"









Db 1157 CCGCTGCATGCCATCATCAACGAGAGACCCTAATGCCGCTGATTAGAGACTGCAGGA 1216  
 Qy 1137 ggaactgcgcagctggaggaacaact-----ccagaagcagtggt 1176  
 Db 1217 GGAAGTGGCCCGCTGGGGAACTGCTGATGCTCAGGAGACTGCAGAGCTTCTGCTTGGGA 1276  
 Qy 1177 ggaagtggaagtggtgaggaaggttct-----gagcgccaaatggaaggaactgtaacc 1229  
 Db 1277 AGGCTTGAAGACGAG 1336  
 Qy 1230 gccgcgaacgcgcgtctgaggaagcaatcgtgtcattcagcagccgagatgagcaagtc-- 1287  
 Db 1337 AGCTCCAGTTTCACCCCTATCCACCACACATTAATGGAGGTGGAGCCGCTATTCTC 1396  
 Qy 1288 -----aagaaatgagcaagcagaagtcgtggaactctggaactcagaggaagtgta 1334  
 Db 1397 CCCCACACGAGAGTCCAGATGTTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1456  
 Qy 1335 gaagcttatacggatctcaatcagacctgggaaggaagcctggcgaagccgaggaat 1394  
 Db 1457 GAAGATTATAGCTGAGCTGAGAGAGAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1516  
 Qy 1395 caaagaaggaaggaagcgcgcgtcgaagagctgtgtatcagcacc-----gaaagag 1448  
 Db 1517 GAGGATGAG 1576  
 Qy 1449 ctgtgtggcccttaccacacccaagaagaatgcaacatagctagctagctgagcagtgatcc 1508  
 Db 1577 AACGTGGGCTCTTCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1636  
 Qy 1509 tctctggctgagtgcttctgtctcaacatcaacacccgagcagaagagctgtggaacgt 1568  
 Db 1637 TCTGATCTGAGTGTGCTGCTACACATCAAGAGATGGCTCACAGAGGCTGGCGAAGT 1696  
 Qy 1569 caaccaagatatacaaacgcggaatctgtctgaagcgttcgaagatc---ctgaaagaaca 1625  
 Db 1697 AGATATGAGATCAAGCTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1756  
 Qy 1626 ctgtgagcttgaagaatgtgacaacgtgtgacacatcgtccaaagcagaagcgtgtctgt 1685  
 Db 1757 CATGCCCAAGCAGATGAG 1816  
 Qy 1686 catgtgtgaacggtgtggaatcgaacgaagcactcgcctccgagcagcgtcagaagatcat 1745  
 Db 1817 ATATGATTAATGGAGAGCTGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1876  
 Qy 1746 ccgtggagattcgaacatttctgattcaaccatccggaagaagcgtgtggaagcga 1805  
 Db 1877 GATGGGCAAGAACCAAGCTTTCCGCTTCACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1936  
 Qy 1806 a 1806  
 Db 1937 A 1937  
 RESULT 9  
 AB014606 6217 bp mRNA linear PRI 06-FEB-1999  
 LOCUS Homo sapiens mRNA for KIAA0706 protein, complete cds.  
 DEFINITION AB014606  
 ACCESSION AB014606.1 GI:3327225  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens adult male brain cDNA to mRNA, clone\_1lb:pbluescriptII  
 SK plus clone:HG04286.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,  
 Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba  
 252-0812, Japan (E-mail:cdna@nifekazusa.or.jp, Tel: +81-438-52-3913,  
 Fax: +81-438-52-3914)

REFERENCE 2 (sites)  
 AUTHORS Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,  
 Kotani,H., Nomura,N. and Ohara,O.  
 TITLE Prediction of the coding sequences of unidentified human genes. X.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro  
 JOURNAL DNA Res. 5 (3), 169-176 (1998)  
 MEDLINE 98403880  
 FEATURES  
 SOURCE  
 location/Qualifiers  
 1..6217  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HG04286"  
 /sex="male"  
 /tissue\_type="brain"  
 /clone\_1lb="pbluescriptII SK plus"  
 /dev\_stage="adult"  
 368..3679  
 /gene="KIAA0706"  
 368..3679  
 /gene="KIAA0706"  
 CDS  
 /note="HG04286 cDNA clone for KIAA0706 has a 23-bp  
 insertion at position 2033 of the sequence of KIAA0706."  
 /codon\_start=1  
 /product="KIAA0706 protein"  
 /protein\_id="PAA31661.1"  
 /db\_xref="GI:3327226"  
 /translation="MAGASVAVAVRPPNARETSODAKCVSMQNTTSITINPKSG  
 DAPKFTFDYSYMSHTSTEDPEQASQOQVYRDIDEEMLIHAFEGYNVICIRAYGGTGG  
 KSYTMGROEGQOQIVPOLCEDLSEFVSESOALSVEYSMEYICEVRLDLP  
 KRSGLRREHPIGLPYODLSKLAVSYADIALMCCGKARFVATNNMTSSRH  
 AVETIVFQRCCHDQLTGLDSEKVKSLVLDLAGEERDSSGARBRKKEGINIKSLT  
 TIGKIVTSALADMSKKRSPDIYPRDSVLTKMLKENGCSRTMTALSPADINYE  
 TLTSLRYADRTKQTRCAITINEDPNALIRELOEYVRLRELMAGLSASALELKT  
 EEGSVRGALPRAVSSPPAPVSPSTINGELEPFSPTESQIGPEBAMRLQETKRI  
 IAEINTEWELKRLTEALRMEREALMEGVAVVEDGTVGVESPKRTPHLVINEDP  
 LMSCELIYHKIDGTVRVGQVMDIKLGOFLREQHCLFRSIPQDGVVTVLECEGA  
 ETVYNGKLTIEPEVLKSGNRIVMKNVFRNHEEADLLEORLADSDSGDSK  
 NFAOKELLEQOQIDIKLEMRRLQDLLENVRAKKEEADLLEORLADSDSGDSK  
 RSCRESRLISSLEBOIPLPTVOTIVRCGLPSGKRRAPRYQIQRRLQKDR  
 WATMADLMQAVKEICTEVALADIRHRAEIEALALMKRELCTYKIPGPGDAMPA  
 VARDVPTVEEESGAGSGGSEEGAEVEDLRAHIDKLTGILQEVKIQNSKDR  
 ELQALRDRLMERVPIPLADHEDNEEGGEVPVAPPEGSAAEAPASDPMPSARP  
 SPPLSWEVRSRLMEEDPAFRGRRLRYLKKOBLQLOGSGGGLRPPRAFEVPP  
 HDCKLRPPKSNPQHRRESWPGSGEAPPLQDPEEYTPHPATPARPPSR8HHNR  
 RNSIDGGRRRGASAPRQHPKKNHSIPQRPYRQRPGRPRYRPTTPRRAR  
 RQNSAPDLKESGAV"  
 BASE COUNT 1406 a 1723 c 1892 g 1196 t  
 ORIGIN  
 Query Match 16.2%; Score 382.2; DB 9; Length 6217;  
 Best Local Similarity 55.0%; Pred. No. 1.7e-97;  
 Matches 1024; Conservative 0; Mismatches 738; Indels 99; Gaps 10;



```
/protein_id="CAA04248.1"
/db_xref="gi:2370435"
/db_xref="SPTREMBL:O35787"
/translation="MAGASVAVYAVRVPFNARETSQAKCVYSQGNSTTIIINPKQR
MFLNASEDYTSWHSITSVEDPQFASQOQVYTRIDIGEMLLHAEGNVCFFAQQGAGK
SYTMNGRQEPQOQIVPOLCEDFSRVNVNOSADLSYSEVSEYSEICERDLINPK
SRGSLRVREHPIGLPYVDLSKLAFTVSYADLADMDGCKNRATVAATMNTSSSHA
VFPIVFORSHDPLTGLDSEKYSKISLVNLAGSRDSSGKRGALRGAINKSLTTF
LGVYSLADLQSKRSDFIPYRDSVYTWLTKNIGNSSTNTAIALSPADINYEET
LSTLRVADRTKOIRCNVYNEDNPANRLIRELQEEVARELRELLMAGSASLGLKVE
EGSPGCVTPAASSPPAPASPSPPPHNSELSPSSPSLEPIGPEEMERLQETEKII
AELNWEKRLKTEALMERELALLEMGSFGWRTGVSPKTPHVLINEDPLMS
ECLLYHKDVTAVGOVDVLIKLGQFIREQHCLFRSIPQDGEVMVTLIEPCEAEFY
VNGKIVTEPVLKSGNRIVMGKNVFRFNHPEORLEBERGVPPEPSPSPVWNFA
OKDWLEQOQIDIKLEMERLQDLENOYRKEKEADLLLEOQLYADSDSGSDRSC
EESWRILSLDELPMYVOTIVKRCGLPSGKRARPRVYQITQRRRLQKDRMAT
MADLKMAVKEICEVALADPRHRAETELAAIKMRELCTYIGKPEPGDAMVAVAR
DVMPTVVEBECGGGGGGEGGARGAEVEDIRAHIDIKLGLQEVKLNSSKRELQ
LRDMLLMERYIPLTODLEDNESGLVTAMPDGSVAVEAVSNDSPAVPSSPQ
SSMWRVSLPMHEDEPAFRGRRLMLKQBLRLOGQSGRGGLRRPARVPDHCK
LSRPFKSNPOHRESMPGSGEAPGPQPEVTAAPPNNRPPSPRRPRHRRNSID
GSSRSGSGSTQPEPQHLRPQKINSYPOQPPYAQNRGPRYPTTTPRRRORSAAP
DLKESGAAY"
BASE COUNT      956 a      1101 c      1269 g      791 t
ORIGIN
Query Match      15.4%: Score 361.8; DB 10; Length 4117;
Best Local Similarity 54.5%: Pred. No. 1.1e-91;
Matches 891; Conservative 0; Mismatches 687; Indels 57; Gaps 6;
QY 219 gaatgtcccaactatgcygacaggaagacctatcccaagatcgcgagtcctct 278
DB 442 GGAGAGCCACAGTTTGATCATCAACAGCAGGTGATGACAGCATAGAGAGAGATCT 501
QY 279 ggataatcatccaagggttaataattgtatcttgcctcaagtcagaccggttcgg 338
DB 502 GCTGATGCTTTGAAAGGCTACAAAGCTGTGATCTTTGCTTACGTCAGCGGGGGCTGG 561
QY 339 caagtcctatcaatgatgagg-----ctatgcycaaggagcatgcygagcccgagat 392
DB 562 CAAGCTCAACACATGATGAGGGGGCGGAGAACCGGGGAGAGGACATTTGACTCAACT 621
QY 393 ttgcgcggaactgttcgcggtatataatgaaactgcagaaggaacaaactcaattgac 452
DB 622 CTGGGAGGACCTTCTCTCGAGTGAAG---TGAACCAAGAGTCTCAACTTCTTACTC 678
QY 453 cgtcgaaatttcgtaacttgaaattcaatgaacgagtgagaaactgtcgaaatcgctc 512
DB 679 TGTGAGAGTGAAGTATATGAGATCTATTTGCAAGAGATGAGAGACCTTTGAACCCCA 738
QY 513 gacaaagggaatctcaaggtcgcagaacacccgctgcagcccgccctacgtgagagact 572
DB 739 GAGTCGGGGGCTCTGCGGGGTCGGGAGACCCCATCTCGGCCCATATGTAACAAGACT 798
QY 573 ggcgaagctgtcgtgcatcatcccaagaataatcgaatgatgaggggcaaca 632
DB 799 GTCTAAATGTGGCTGTGACTTCTCTATGACAGATGCTGACTGATGACTGTGAAATAA 858
QY 633 agccagaagcgttgcgcgcacaaacatgaagagagacatcagtcgaccgcgctct 692
DB 859 GGCAAGAACTGTGCTGCTGCCACCAACATGAAGAAACAGACGCGCTGCCACGCTGCTT 918
QY 693 caacttgaacttgcgcgcaaaagtgcatalgaaagagacacaaatggacacagaagaagt 752
DB 919 TACTATGCTGTTTACCCAGCGCTCCCATGACACACTTACTGATGGATTGAGAAAAGGT 978
QY 753 tgcgaagataagtcgtgtagattggcgggttcgtcagcgaacacgtcaccggaagtac 812
DB 979 CAGTAAATCAGCTTGGTGAACCTTCCGGGAGGACGCGGCGACATCTCAGGGGCTCG 1038
QY 813 tggagcgagctgaagagaggtgacagatcaaccgctcaacttcgaccctagtcggt 872
```

```
DB 1039 GGGCATGCGCTCTGAAGGAGAGGTGCAAAATCATATAGTCCTGACTACTGTAGGAGGT 1098
QY 873 gattgcaagcgttagcgaatgltcgtcgcgaaanaaagaagaagaatcagtlagtaactta 932
DB 1099 GATCTACGCCCTTGCGAGATTTGCAATC---AAGAAGGAGGAAGTGGAGCTTCACTCCCTTA 1155
QY 933 ccgaatcctgtaactggaagtcgtctctctgaaggactcccttggaggcaactcgaatgaccgc 992
DB 1156 CAGAACTCTGTGCTTACTCTGCTACTCAAGAGATTTGGGTGGAACTCAGCACAGC 1215
QY 993 catgattgcgcgcatlccgcgcgtcgatalttaacttgaagagactccagtlacccttcg 1052
DB 1216 AATGATTCAGAGCCCTCGAATCCCGCTGACATCAATTACAGAGACACTCAGACTCTCAG 1275
QY 1033 atatgccaactctgcgaagcgaatcaagaacacagcagtggtlcaatgaagaacccgaacgc 1112
DB 1276 GTACGCAACGCCCAACCAACAGATCCGATGCAATGCTGTCAATCATGAGACCCCAATGC 1335
QY 1113 ggcggtatcgcgaggttgaagagaggaactcgcgagctgagggcaact----- 1163
DB 1336 CCGGCTTATCCGGGAGCTGCAAGAGAGAGTGCCCGGTTGCGGAAACTGCTGATGCTCA 1395
QY 1164 -----ccagaagcagtggtggaagtggaagtggtgagagaggtctcgcgcgcga 1212
DB 1396 GCGGCTTTCAGCCCTGCTGCTAGAGAGGCTTAAAGGTGGAAGGAGGAGTCTCGAGAGTGT 1455
QY 1213 gtggaagaaatcgttacccgcgcgaacacgcgcgtcgaagaacaaatcgtgcatcaagcag 1272
DB 1456 TCTGCTGCTGCTCATCATCTCCCTGCCCCAGCTTCAACCTCATCTCCTCCACATTA 1515
QY 1273 ccgagatgcagacatcaagaataatgagcaag-----ggaagaat 1310
DB 1516 GCGGAGCTGGAAGCCGCTCATTTTCAACCAAGCTGTGAGCCCCAGATTGAGGCTGAGAGGC 1575
QY 1311 cgtggaagcaacttgaacagagtgagaagctctatccgagatccatcaagactcagagtgaga 1370
DB 1576 CATGAGAGAGCTCGCAGAGAAACAGAGAAATTAATAGCTGAGCTGTAAGTGTGGAGGA 1635
QY 1371 gaacttgcgcgaagccgagaggaatccacaagaagcagagagcgctcgaagagcttgg 1430
DB 1636 GAACTCTCCGTAAAGCAAGAAAGCTGTGAGAGTGAAGAGAAACATGCTGCTGAGAGTGG 1695
QY 1431 tatagcaatcgaaaaggccttctgtgccccttaccactcccaagaatgccaactag 1490
DB 1696 CTCGCCGAGAGAGTGGGTACTGTGGGTCTTCTCTCCAAAAAGACTCCCACTGCTGT 1755
QY 1491 caacttgaagcgaatgactcctctctgtgagtggtctgtgtacaaactcaagcccgagca 1550
DB 1756 AAATCTGAATGAGAGACCCCTTGATGCTGTAATGCTTCTACACATCAAAAGATGGCGT 1815
QY 1551 gacaaggttgaagacgtlcaacccaagaataacaaagcggaatcgtgtgaaaggttgcga 1610
DB 1816 CACCAAGGTTGCGCAGGTAGTGTGACATCAATTACTGCGCAGTTCAATCCGGAGCA 1875
QY 1611 gatcctgaagaagaacactgtacgttltgaaa---atgttgacaacgcttltgacaactcgtg 1667
DB 1876 GCACTGCTCTTCCGAGACATCCCTGACGCTGAGTGAAGAGATGATGTGATGATGAAAC 1935
QY 1668 aaacgagaagagctgtcgtcatgltgaaagcggtggaatcgaacaagcctaactcgctcgc 1727
DB 1936 TTGTGAAGGAGCTGAGACATATGTCATGGAAGCTTGTGACGAGCGCGCTGGTGTGAA 1995
QY 1728 cagcggttaagagatccctcgtgagatccacatlltccgatataacacatccogaaga 1787
DB 1996 GTACGGAACAGGATTTAATGGGCAAGAACCAATGTGTTCCGCTTCAATCAACCCGGAGCA 2055
QY 1788 agctcgtcgaagc 1802
DB 2056 AGCACGCGCTGGAAGC 2070
RESULT 11
AB074017
```

LOCUS	AB074017	3303 bp	mRNA	linear	ROD 16-JAN-2002
DEFINITION	Mus musculus klf1c mRNA for kinesin superfamily protein 1C,				
ACCESSION	AB074017				
VERSION	AB074017.1	GI:18181920			
KEYWORDS					
ORGANISM	Mus musculus				
SOURCE	Mus musculus cDNA to mRNA.				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
TITLE	1 (sites)				
JOURNAL	Nakajima, K., Takei, Y., Tanaka, Y., Nakagawa, T., Nakata, T., Noda, Y., Setou, M., and Hirokawa, N.				
MEDLINE	Molecular Motor KIF1C Is Not Essential for Mouse Survival and Motor-Dependent Retrograde Golgi Apparatus-to-Endoplasmic Reticulum Transport				
REFERENCE	Mol. Cell. Biol. 22 (3), 866-873 (2002)				
AUTHORS	2 (bases 1 to 3303)				
TITLE	Nakajima, K. and Hirokawa, N.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (08-NOV-2001) Kazuo Nakajima, University of Tokyo, Department of Cell Biology and Anatomy, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail: kazuoem-u-tokyo.ac.jp, Tel:81-3-5841-35841, Fax:81-3-5683-4856)				
FEATURES	Location/Qualifiers				
source	1..3303				
gene	/organism="Mus musculus"				
CDS	/db_xref="taxon:10090"				
	1..3303				
	/gene="Klf1c"				
	1..3303				
	/gene="Klf1c"				
	/codon_start=1				
	/product="kinesin superfamily protein 1C"				
	/protein_id="PAB83862.1"				
	/db_xref="GI:18181921"				
	/translation="MAGASVAVAVRPFARETSDQACVSNQMGTTIINPKOSK DAPKSTFDYSWHTSEDEDFEASQOOVVRDIEEMILHFEEDYANCIFAYGGTAGG KSYVMGROBPQOQIVPOLCEDFESVYVADNIOSSVSYVMEIACEVRLLNN KSRSLRREHPILGPYVDLSTKAVSYADIDLMCGNARFVAVDANNETSRSRSH APEFIVTQRSHOILNGLDSKPKVSLINDLAGEBNDSCARBNKKEGNIKSLTF TGVFISALADLDSKKRKSDFPIRDSVLYTLWLEKNGNSRTMIALNSPADINTEB TSLTLYRADRTQIRCNAINEDNENARILRLQLEVARLRDLAOGISASALGILVW EEGSPGGVLPASSPPAPSPSPPPNGELPEPSSAEPIQIEEMERLQETEKI IAEINTEMEKRLKTEALRMERALLAMEGVAVEDGGTVGFSPKTPHIVNEDPQ LMSCLLYHKIDGTVRGOVDVQIKLGOFFIREOHCFRSTIPDQGEVMTLPECEG ETVYNGKLTELRLVLYKSGNRYVMKKNIVGKIRFVFNHREBQARLBERGVRPPRPPSPSDV MFAOKELBOGDIDIKLEMERLQDLENOIKKEEADLLLEORLTADSDGEDSDV RSCSEWRLLISLREQLPPTTVQNIIVKRGLEPSSGKRRAPIRVQIIPQRRLQKDRP WATVADLKQAVKEICYEVALADIRHAELEALAAAKMRKLCITCYGPEGBQAMRAA VARVMDVTVEEGGGGGGSEEGAGAEVEDIRAHIDKLTGLQEKLONSSKDEBE LQALRDLRMERIVPTLPDLEDNDSEGLVTMAPPGSPAVEETVNDSDPAYRPS PRLSSRSVSRILMEEDPAPFRGRRLRWIKODRLRQLOAGAGRGGLRRPARVPPRPP DCKLRFPFKSPHQRRESMPGSGEAPAPPEPEVTPAPAPNKRPPSPKPRSRKRR SLDSGSRSGGGSTQPEPQRLRPQKHNGYPOQPPSPAQRGPYPPYTPPRMRORQ SAPDKESGAAY"				
BASE COUNT	757 a 930 c 1013 g 603 t				
ORIGIN					
Query Match	15.4%; Score 361.6; DB 10; Length 3303;				
Best Local Similarity	54.5%; Pred. No. 1.2e-91;				
Matches 936; Conservative 0; Mismatches 714; Indels 66; Gaps 8;					
151	agtgcanaactatgatgagcccggaagcattgcttcgatcgatcgatgctc- 209				
115	AATCCCAACAAGACGACGAGATGCCCCCAAAAGCCTTCACCTTATCTTACGTCG 174				
210	-cttcgaagaacgctcccaactatcgagacagaagaccttccaagatctcga 267				
175	CATACCTGGGTGGAGATCCGACAGTTTGCATCTACGACGACAGGTACCGACATAGGA 234				

OY	268	gtccgcgtctcgtgataatgcatcacaagggtltaacaattgatacttcgcgtacggtcag	327
Db	235	GAAGAGATGCTACTCCATCGCTTTGAAGGCTCAATAGTGTGCATCTTGCTACGGGAG	294
OY	328	accggttcgagcaagttcctattcaatgaggg-----ctatggcaaggagcatggcgtg	381
Db	295	ACGGGGGCTGGGAAGTCTCAACCATGATGAGGGGGGCGACAGACCCGGGACACAGGCAT	354
OY	382	atccgcggaatttcgcagagcatgcttcgagcgatataagcatcgaatcgaagaagaac	441
Db	355	GTGGCTCAGCTCTGTGAGGACCTCTTCTCTCAGTCAATGTAACCAAGTGTCCAG---	411
OY	442	ctcaactgcaccgyltcgaagtttcgtaacttggaaatttaacatlgaaagagtgcgagactg	501
Db	412	CTTCTCATTCAGTAGAGGTGAGACTCTATATGGAATCTATTGGCAGAGTACGAGACCTC	471
OY	502	ctgatactcgttcgaacaaggaggaatctcaaggtgtccgaagacaccggtgcacggccctac	561
Db	472	TTTGAACCCCAAGACTCGGGGCTCTGTGCGGGTCCGGGACACCCATCTTAgGCCGTGAC	531
OY	562	gtggaggaagcttgcgaagcgtgcgttcgtgcgatactcacaagaatcgaaaatctcatgat	621
Db	532	GTGCAGAGCTCGTCTTAACATGCGGTGTGACTCTCTATGACGACATTTGCCGACTCATGAC	591
OY	622	gagggcacaaagccagacaggtgtccgcgcaacaatgaaacagagacatccagttgcatacc	681
Db	592	TGTGGAATAATAGGCANAGAACCGTGGCTGCACCACCATATGACGAGACAGCGCGTCC	651
OY	682	caagcgcgtttcactcttgtaacttgcacgcgcaaaagtggcatgatagaaagagaccaaattgac	741
Db	652	CATGCCGCTTACTATTGTTGCTTTTACCACAGCATCCCATGACACTTACTGCGCTGGAT	711
OY	742	acaagagaagttcgcgaagatcaatctgataagatttgcggtgttcgcagagagaacgtcc	801
Db	712	TCAAGAAAAGTCACTAATAGATCAGCTTGTTGTGACTGTCCGGGAGTGAAGCGGCTGACTCC	771
OY	802	accggaagctacttgcagcgcgactgaaggaggtgtcagagatcaaacgcgtcacttgcacc	861
Db	772	TCAAGGGGCCCGGGGATGCGCTCTTAAGGAAGCGCCACATCAATTAAGTCCCTGACTACG	831
OY	862	cttagtctgigtgattgcagagcgcttagaggaatgactgtcggagaaacagaagaagaatacag	921
Db	832	CTTGGGAAGGTGATCTTAGGCCCTGTGGCAGATTGCAATC---AAAGAGGGGAAAGTCGGAT	888
OY	922	ttaagtacctaacccagaattcgtgacttgcagtggtctcttgaagaagactccttgcggaagcac	981
Db	889	TTTATCCCTATATAGACTAGTCACTGCTTACTGTGCTCTCAAGAGAAATTTGGTGTGGAAC	948
OY	982	tcatgtagcgcgaatgattcgcgcgaacttcgcgtctgatatataacttgaagaagactctc	1041
Db	949	TCTGTGCACAGCAATGATTGGACGCCCTGTAGTCTCGTGCATCATTAATACGAGAGACACTTC	1008
OY	1042	agtatcccttgataatgtcgagcctctgcgaagcgaaatcaagaacacagcgagtgtgtaatgaa	1101
Db	1009	AGCACTTCACAGTATGTGACAGCCCGCACCAACACATCCGATGCAATGCTGTCTCAACAG	1068
OY	1102	gaaccgaacgcgagcatcgtatctgcgcgagattgaaggagaaactcgcgcagcttgcggagcaaa	1161
Db	1069	GACCCCAATATGGGCGCTTATTCGGGAGACTCAGAGAAAGAGTGTGCCGCTGCGGAGACTGT	1128
OY	1162	ctcgaagagcagttgtgtgaggtgtgaggtgtgtcagagaagttcttcgcyggscagttgag--	1218
Db	1129	CTGATGGCTCAGGGGCTGTGACGCTTGCTGTGAGAGGTCTTAAGAGTGAAGAGGGAGT	1188
OY	1219	-----gtaatcgttaaccgcgccgaagacgcgcgtctgagaagaac	1254
Db	1189	CTGTGAGGTGTTTGTGCCCCCTGTGCATCATCTCCGCCGCCGACGCTTACACCTCATCTCT	1248
OY	1255	atcgtctgattcagcgcgcggatgtcgaacgtcaagaanaatgacgaagcagaatcg--	1312
Db	1249	CCCCCAATATATGGGAGACTGTGAGGCCCTCATTTCTACCCAGTGTCCGAGCCCCAAATTTGG	1308
OY	1313	-----tgagaagcaatcgaaccagagtgtgaagcgtcatatcgtgattcacaacag	1359

Db	1309	CTGTAGAGAGCCATGGAGAGGCTGCAGAGAAACAGAGAAATCATATAGCTGAAGTCATATAG	1368
QY	1360	accttggagaagaaagcttggccaagaacggaagaaatccaaagaaagaaagcgagctc	1419
Db	1369	ACTTGGAGAGAGAACTTTCGAAAGACAGAGAGCGCTGGAATATGAGAGAGAAAGCATTGCTG	1428
QY	1420	gagagagcgagg-----atacgaatcgaaaaagggcttgttggccctaccacatccaaa	1473
Db	1429	GCTGAGATGGGCGTGGCTGTCCGGAGAGATGGCGGTACTGTGGGTGCTCTCTCCAAAA	1488
QY	1474	gaaatgccacatctagtaacttgaacttggacgatcatccctcttcggctgaagtcttctac	1533
Db	1489	AGACTCCCGACCTGTAATCTGAAATAGAGACCCCTTGATGTGTGAATGTCTTCTCTAC	1548
QY	1534	aacatcaagccccggcgagacaagggcttggaaacgttcaaaccaagatatacaagcggaaat	1593
Db	1549	CACATCAAAAGATGGCGGTACACAGCGTTGGCCAAATGATGAGACATCAAGTTGACTCGC	1608
QY	1594	cgctggaacggttgcgaagatccctgaaagaacacgttaagtttgaatattggaacggt	1653
Db	1609	CAGTTTCATCCGGAGACGAGATTGGCTCTTCCGACACATCCCTCAGCTGATGAGAAATG	1668
QY	1654	gtg---accatcggtcccaacagagaagcgctgctgtcatagtgaaacgagctgcgaatcac	1710
Db	1669	ANGGTACTCTGTGAGCGCTTGTAAGAGAGCTGAGCATATATGTCATATGGGAAGCTGTGACG	1728
QY	1711	aagcttactcgctccgcagcgcgctcacaggtatcatccctggggcgatcttcacattttcga	1770
Db	1729	GAGCTGCTGCTGCTGAAAGTCAAGAAACAGGAATGTAATGGGCAAAACCATGTTTCCGC	1788
QY	1771	ttcaaccatcccggaagaagctcgctcggaagcgcaa	1806
Db	1789	TTCATTCACCCGAGAGAGCGCGCGCTGGAACAGAA	1824

RESULT 12

CEL0104A

LOCUS

DEFINITION

Caenorhabditis elegans kinesin-related protein (unc-104) mRNA, complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

Location/Qualifiers

1..5170

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

1..5170

/partial

/gene="unc-104"

1..5170

/gene="unc-104"

111..4865

/gene="unc-104"

/standard\_name="uncoordinated-104"

/note="putative; putative"

/codon\_start=1

/product="kinesin-related protein"

/protein\_id="AA03517.1"

/db\_xref="gi:156478"

/translation="MSSYKVAVVRPFENQREISNTSKCVLQVNGNTTINSHINKEN

Query Match	15.0%; Score 352; DB 3; Length 5170;
Best Local Similarity	54.8%; Pred. No. 7,1e-89;
Matches 918; Conservative	0; Mismatches 680; Indels 78; Gaps 8;
QY 133 gaagagaagcgcgctaaagaatgycgaactatitagtatgagcccggaagcattgcgtc	192
DB 198 GGAAATACGACGACATATAATAGTCATTCAATTAACMAAGAACCTTCAGTTCAATTTT	257
QY 193 gatcgcgtatctgctccttgacaagaatgctcccaactatgcgacagaagaaccta	252
DB 258 GATCATCTCATATTGGTCGTTTGGGAGGATGACCACATTATTATACCCAAAACAGTA	317
QY 253 ttccaagatcgcgagtcgcgcgtctctgtaataatgatacaaggtatacaaatgtalc	312
DB 318 TATGAAGACCTCGAGGTTGAAATGTTGGAAACGCGATTGAAGGGGTATATGTCGCATT	377
QY 313 ttgcgtacgttcaagaacggttcgcggaactctcttaataatgagtcataagaaag	372
DB 378 TTTGCTACGAGTCAACACAGATCAGGAATATATATACATATATGGGAAACCCATGAT	437
QY 373 c-----atgycgtgtatcccgcggaattgcgcggaatgtttccgcgctataatgaa	423
DB 438 CCAGATGAATGGGCTATATTCACGCTTGTGCATGATTTATTGACAGATTGATAA-	496
QY 424 ctgcgagaagaagaactcaacttcgcacgcgcgaagtttgcctacttgtaaatatacat	483
DB 497 --TAAACAATGATTAAGATGTTCAATATTTCTGTAGAGGTAATCTATATGGAATTAATGCT	554
QY 484 gaacagtgtcggaacttgcttgtaactcgcgtcgaacaaggggaactcaaggtccgagaacac	543
DB 555 GAACGAGTAAAGATCTTGTGAATCCTTAACCTCGGAGGTAACCTTAAGGTTGCTGTAACAT	614
QY 544 ccgtgcacgcgcctcaactgctgtaggaacttgcgcgaagctgctgtcgatcatctcaagaa	603
DB 615 CCTTCTCTTGAGACCTTACCTGCATGACCTTACCAAAATGGCAGTTGTTCTTACACGAC	674
QY 604 atcgaataatctatgtatgaggaacaagaagccaagaagttgcccgcacaacaatgaac	663
DB 675 ATTTCGAATCTGATGACGACGAGAAATTAACGAGAACCTGTTCTGCCAACAAATATGAAAT	734
QY 664 gaaacatacagtcagtcaccacgcgctcttcaactttgacattgcgcaaaagtgcatagat	723

Dh 735 TCGACATCATCAGATCCCATGCGGATTTACACATATGACTCATCTCAAAAAAGACATTGT 794  
Qy 724 gaagagacaaaatgacacagaagaaggttcgaaagatcagtcgtgatttcgaggt 783  
Dh 795 GCTGATTTTAATTTGGATACACGAGACATCAAAATTTCTTTGGTTGATTGGCAGGA 854  
Qy 784 tctgagcagagaacgctcagcagagctactcgaagcagactcgaagaggttcagagatc 843  
Dh 855 TCCGAAAGACCAATTTCTACAGAGACGGAAGTCAACGACCTAAAGAGAGCAATATTC 914  
Qy 844 aacgcctacttcgacccctagtcgtgatttcgagcgtcagcagatagtcgtc---g 900  
Dh 915 AACCAAGACTTGGACAACTTGGGCTTTGTAATCACTAACTTGAAGAAATCAACCAAG 974  
Qy 901 ggaagaacagaagaacacagctagctacccatccagagattcgtctcagtcgtgcttc 960  
Dh 975 AAGAAAGTCCAAACAAAGGTGTGATTTCTTATCTGATTTCTGTCGATCGGCTTCG 1034  
Qy 961 aaggaactccttcgagagcgaactcgaatgcgcacatgatttcgcgcatttcgccttc 1020  
Dh 1035 AGAGAAATCTCGGAGGAAATTCGAAACTCGCATGCTCGCGCATATTTCTCCACTGAT 1094  
Qy 1021 attacttcgagagacactcgaatccttcgattcgtatgagagctcgcagagagatcaag 1080  
Dh 1095 ATTACTTTGATGAAACCTGAGTACTTTGAGATATGCGGATAGAGCGAAACAAATTTGTT 1154  
Qy 1081 aaccacgagctgctcaatcgaagaaccgagcgcgagatcgcgcagcttcgagagagaa 1140  
Dh 1155 TGTCAAAGCATGTTGTTAATGAGAGATCCAAATGCAAAACTTATTCGTAACATCAAGAA 1214  
Qy 1141 ctgcgcagcctcgaagagcaaacctccagagcagtcgagtcgaagtcgagtcgagagtc 1200  
Dh 1215 GTTATCAAACTTCGACATTTTGAAGATTAAGG----- 1249  
Qy 1201 tctgcgcgcgcagtcgaggaatcgtacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1260  
Dh 1250 -----AATTGATGCTCAGACAGTACGAGGAGACCCGGAACACAAAAAGAA 1298  
Qy 1261 tgcattcagcagccgagatcgcagcagctcagaagaatcgaagaagcagaatcgttcgagca 1320  
Dh 1299 CCGAAATTTGCCAGC-----TCATGTATCATGAGAAC-----TTGAGAA 1337  
Qy 1321 ctgaaccagagtcgaagaactctactcggagatcgaatcgaactcgcgcgcgcgcgcgcgc 1380  
Dh 1338 CTTCAGAGATTCGAGAAATTTGATGCTGCTGAAATTTGAAAGACTTGGAGAGCAAAATTAAT 1397  
Qy 1381 aagacgcgagaaatccacaagaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1435  
Dh 1398 CATACAGAAAGAAATTAAGAAACACGAGAGGAAAGAACTTCGTGATATGGGCTTTGCGTGT 1457  
Qy 1436 gcatcgaagaagggcttcgttcgacctaccactccaaaga-aatgcacacatcagtcacac 1494  
Dh 1458 GCAGAAAGATGGAACAACTTAGGCGTATTTAGTCCAAAGAACTTACCCATTTAGTTAAC 1517  
Qy 1495 ttgagcgtatgacctccttcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1554  
Dh 1518 TTGAATGAGATTCATATGTCAGAGTGTCTCATATATATTAATTAAGAAAGAGCTAACT 1577  
Qy 1555 agggcttcgaacacgtcaacaagaatacacagaagcgaatcgtcgtcgaagctcgaagatc 1614  
Dh 1578 AGCGTGGGCGCCCGCGAAGCTGACATCGTCCAGACATTTCTTTATCGGGAAGACGATTT 1637  
Qy 1615 ctgaagaacacacgtcagcttcgaaaaatcgtgcaaacgttcgttcgacacacgag 1674  
Dh 1638 CTGCAATTAATTTGATTTTATTAACGAAGATGGAATGACATATGCAATGAAACCG 1697  
Qy 1675 aagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1734  
Dh 1698 AATGATCTTTGTTATTAATTAATGAAACAAAGTACAACTCTCATGTTATACACACAGGA 1757  
Qy 1735 taccagatcactcgtcgc 1790  
Dh 1758 TCTAAGTATTTCTCGGTGATCAATCAGTTTCCGATATATATGATTCACAGGAAGC 1813

RESULT 13  
Dh081788  
LOCUS  
DEFINITION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 7107)  
Li, H.P., Liu, Z.M. and Nirenberg, M.  
Proc. Natl. Acad. Sci. U.S.A. 94 (4), 1086-1091 (1997)  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
2 (bases 1 to 7107)  
Li, H.P., Liu, Z.M. and Nirenberg, M.  
Direct Submission  
Submitted (10-DEC-1996) Laboratory of Biochemical Genetics,  
National Heart, Lung, and Blood Institute, Building 36, Room 1C-06,  
36 Convent Drive, MSC 4036, Bethesda, MD 20892, USA  
FEATURES  
source  
1..7107  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
299..6064  
/note="Kinesin heavy chain family member"  
/product="Kinesin-73"  
/protein\_id="AAB50404.1"  
/db\_xref="GI:1906596".  
translation="MASDKIVAVRVRPPNREIFLDTKCIYVEKQOTLIQNPPL  
KIERKQPTFAFDHCFYSLNPEDENFAQETVFDVGGLIDNAOGNACIFAQGT  
GSGKSYTMKGQESKGIIPRLCDLFSIAKNSPELMTKVEYSMEYKNEVHDL  
PKPNQSLKVEHNMGYPYDGLSOLATYSODIDNLTGKSKSTVATATNMNAESR  
SHAVSVYLTOITDQATVGSEKYSRMSLVDLAGSEAVATGAGDMLKGSNKS  
LTGLIVSTKLADONGKSGNDKVPVRSVLTPLKLDNIGKSNRYTVAATISAD  
NYEETLSTRADRAKRIYNNVAVNEDPNARIIRLRHEVELTKSLMHLNHAIVPGDV  
QDKLAESNLNQISQTEKLVKTERIQNERQQLKEMGISVQASGKIVKKNKYLV  
NDMDPSLNELLYLLKDRTLIGRTIGSQDPDIDLSGLIOPEHCVTIIEGSLYME  
PVQGRFCVNSAAVEKPIQNGDRIILGNHFFVNPKNSTMCASEPOTPAOL  
DYNFARDIMONEISNDPIQTAIARLEOHEEDKQVATLEKOEVEKQFOOLRNLIS  
STPYAPVAPYDPLBNGKITTPNPTSSOMVEKMAQDRDMPRSLSGLKTDIMRANSLV  
OENFLAEMEKRTFSTTLQIPPNLSPNRRGATFSEPALAKRTISGSOIWMK  
LENKLTDRREMYQEHKEVNLGLPLIEPSEDEFDDKEDNAKPPDPRYESQENMLI  
GVANITFLEFHDVRLDYHTPIISQCGVAVRLOVEIRIAGOMPODMCESVSSG  
DSRDEYDVPDPTSMQITCRVYIKCASGLPLSLNFCQYTFMGHOMVAVPINAES  
TAHDNMVKEFEHQDDFTVTINEELFICIGALSIEWNGHSAFESTKIGVEVOOO  
AKARSLVDMELSKIELWEIHELNGEVSPEVNNISLVAGGITYOLRQGOOR  
VNVRRPYONGSTPLIICQSTYNVAGSVYVRSRQRPLDSIQEEDLVLEKSEAL  
GRROYLDOOTOMLTKKEKNEBERELNSHOWSLTEENALVLPAPSGTGPAG  
ASWEPESGMEHPHVPVLFLLNKGDDLSAQNTDELVSAGINSTLSKHEHKKYTLQILQ  
HLDKQVCCVASMDSHDSQALNRYTEANERYLLRTTYSHPAPMDLVLRKLSI  
NIKKGQTLDRLKRLVRGENAIIWQSEVTEEVYNIKASLELDRSLALASGD  
DCSASDEGTYTEKTRGVSAVESITLRLRONVAVKLETAHQGLSMRTVSPNF  
SOOLINKLTOIMRDPASMESLINVRSRSPDILNLSAANKFTPGHSPAGAGYIRS  
RHSFGKSSSDSPCKARGLNSPATSKLGMATVLEHPLDGHSLDEDEPDSIDS  
EYAAVEQERQONKSMATRSRLTAJRTKSPFADVSISHQSTLSTSSANAMKMLTG  
LATLSMSSSTSGVGSQAVNSCNLSNEDIASRMSIDETPDPFVANSNPNRQARV  
NPFLKDMKRAKIOEOPROAKLOEATFHPLEQLSREMAQSDDECAOLPKNNNV  
DLVNEFKPLSGQMELEFPMQSEOTEFATDNQONRSDLSHSEDLDEGDIYRE  
ELPAAKYVRKKSNTQPPSNKSNINNNNGTQVPRINHRSVAAKBEALAYLDSIM  
TSSTVEDBSKVELVLEPMLVGEVSILRPNTSGVIRVGTGTFQEGAMVGLDT  
PTGKNDGSKVGQYFQCKPKHGMFVRSDKLMDKRGKAMRAYKAERKSNISKEKST  
MTGSMTRKSRSDSLNLSARK"  
BASE COUNT 1942 a 1773 c 1852 g 1540 t  
ORIGIN



Query Match	15.0%	Score 352	DB 3	Length 7107
Best Local Similarity	59.2%	Pred. No. 7.5e-89		
Matches 689	Conservative 0	Mismatches 445	Indels 30	Gaps 4
QY 6	gggcggttgaaatcaagagtggtgtgctgagagccgcttcaacgcgcgaagatcga	65		
DB 301	ggcagatgataagatcaacaaagtcgccgtgaagagtggtgagagagagagagagag	360		
QY 66	ccgttgccgaatgattgtgtgagagaaatcaaacatccctccctcc	125		
DB 361	actagatacgaaatgtatctgtgaaatgaaaaacagagacgatactgcaaatccgcc	420		
QY 126	gggtgcccgaagaagagggcgctaaagatgycgaacatattatgtatgcccgaagcatt	185		
DB 421	gccactggaaaaaaatgagagaaaaa-----ccaaacaaatt	459		
QY 186	tgcgttcgatcgttcgtattgttccttgacaagaatctcccaactatgagagaaga	245		
DB 460	cgcatctcatcactccttttactcatttaaccccgacgacgaaatttgcctccacga	519		
QY 246	agacctatcaagatctcggagtcgcgctcttcgtatataatgacatcaagtgattaca	305		
DB 520	gacatgtgttcgattcgtggcgacgtggaattctggatattgattccagggctattatgc	579		
QY 306	ttgtatcttcgctcaggttcagaccggttcggggaagtccttaatagtatggtctatg	365		
DB 580	gtgcattttcgtttcggccacacaggttttggcaagtcctacagatgatggaccca	639		
QY 366	caagagatgtagcgtgaccccgagatttgcagagacatgttcggcgattatgaact	425		
DB 640	ggagagcaagggcatcattccacgctctgttgaccactctttctggccattagccacaa	699		
QY 426	gcagaaggaacaagaacctcacttgaccggtcgaatttcgtacttggaaattacaatga	485		
DB 700	atccacaccccgag---cttatgtacaaagtgtaggggtgcttccatgagatctatagca	756		
QY 486	acgagtgagagacttgctggaatcc---gtcgcaaaagggaatctcaaggtccgagaaca	542		
DB 757	gaaggtccacgactgctgcgtgaccccaacccgaacaaacagtgcttaaggtccggagca	816		
QY 543	cccgctcgacgcgcctcactggtgagagacttggcgaagctggtgcgtgcatactccaaga	602		
DB 817	caatttttgggcccattatgttgacggactgttcgacgctgctgtgacattcttatcagga	876		
QY 603	aatcgaaaaatctcatgatagtgaagggcaacaagaacccagaaagtgccgcacacaatga	662		
DB 877	catcgataaacctcactgacgaggaagcaacaaagtcagaaagggggccgacaaatgaa	936		
QY 663	cgagacatccagtcgatcccaacgcgcttccacttgcacttgaagcgaagaagtgcata	722		
DB 937	tgccgagtcgtgcgcctccacgcgcgtcttcacatgtgcttccactcactcactcagca	996		
QY 723	tgaagagacaaatgagacacagagaaggttgcgaaatcagtcgtgtatattgacgg	782		
DB 997	tcagccgcacggagtgacggcgcaaaaggtgtcccgatgtcccttggttgatttgactgg	1056		
QY 783	ttcttgagcagcaacgtccacccggagctacttgaagcgcgactgaaggaaggtgacagaat	842		
DB 1057	ctccgaggggggctgtgaaagacggagcgtgttgccgattcttcacaggaagggcttaact	1116		
QY 843	caacgcctcaacttctgaacctagtgcgtgtgatttgcagcgtacgagatatgtcgtcgg	902		
DB 1117	caacaaatcttgacacacacttgggctgtgctcactcacaagttggggcgacaaatccaa	1176		
QY 903	aaaacaga---agaagaatcagttgtatccctacaggaagtttggatgactgagtgctt	959		
DB 1177	caagaaagacggtacgaacaaattggttccctacgcgcgattcgtccatccctggtgtgt	1236		
QY 960	gaagagctccttggaggaacatcgatgacgcgacatgattgcgcgcatctgcgtgta	1019		
DB 1237	gaagacaaatttgggtggcaattccagaaacttttatgtatggagacaattctcccatcgcc	1296		
QY 1020	tattaacttgaagagactctcagtaaccttcgatatatgcgaaactctggaagcgaatcaa	1079		

Db	Accession	Gene	Protein	Length	Score	E-value	Organism	Reference	Authors	Journal	Year	Volume	Page	DOI	PMID
Db	1297	AGATACCTACGAGACGCTTTCCACGCTGCTTATGACAGTCAAGCCCATTTGCT		1356											
Qy	1080	gaaccacgcagctgtgtcaatgaagaccgaacgcgcgcgaatgtaccgcgcgaattgaagagagga		1139											
Db	1357	TAACACGCGCTGTGTTCACGAGAACCCCAATGCCCGCATCTATGTGTAGCGTGGCAGCAGCA		1416											
Qy	1140	actgcgcgcagcttgagaggaact		1153											
Db	1417	GGTGGAGACCTTAGAAGTATGCT		1440											
RESULT	14														
AB037923															
LOCUS	AB037923	6618 bp	mRNA	linear											
DEFINITION	Mus musculus mRNA for KIF13A, complete cds.														
ACCESSION	AB037923														
VERSION	AB037923.1	GI:10697237													
KEYWORDS	KIF13A.														
SOURCE	Mus musculus (strain:ICR) 4 week male brain cDNA to mRNA.														
ORGANISM	Mus musculus														
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.														
AUTHORS	1 (sites)														
TITLE	Nakagawa,T., Setou,M., Seeg,D., Ogasawara,K., Dohmae,N., Takio,K. and Hirokawa,N.														
JOURNAL	A Novel Motor, KIF13A, Transports Mannose-6-Phosphate Receptor to Plasma Membrane through Direct Interaction with Ap-1 Complex														
PUBMED	Cell 103 (4), 569-581 (2000)														
REFERENCE	11106728														
AUTHORS	2 (bases 1 to 6618)														
TITLE	Nakagawa,T. and Hirokawa,N.														
JOURNAL	Submitted (02-FEB-2000) Terunaga Nakagawa, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033 Japan (E-mail:terunagaem.u-tokyo.ac.jp, tel:81-3-5841-3336, Fax:81-3-5689-4856)														
FEATURES	Location/Qualifiers														
source	1..6618														
gene	/organism="Mus musculus"														
CDS	/strain="ICR"														
	/db_xref="taxon:10090"														
	/sex="male"														
	/tissue_type="Brain"														
	/dev_stage="4 week"														
	99..5348														
	/gene="kif13a"														
	99..5348														
	/gene="kif13a"														
	/codon_start=1														
	/product="KIF13A"														
	/protein_id="BAB16346.1"														







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:48:52 ; Search time 378.07 Seconds

(without alignments)  
10681.045 Million cell updates/sec

Title: US-09-235-416-2

Perfect score: 2352  
Sequence: 1 atgtcggcgcgtgtggaatat.....aaacagcgaagcagaattc 2352

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2352	100.0	2352	20	AAx87656
2	495.8	21.1	5322	23	ABL14183
3	473.2	20.1	8309	22	AA159190
4	382.2	16.2	1416	22	AA160976
5	380.6	16.2	3930	21	AAZ44744
6	380.6	16.2	3930	22	AAAD08139
7	352	15.0	7134	23	ABL12291
8	321	13.6	2275	22	AA531140
9	316.2	13.4	10682	21	AA65550

10	309	13.1	1701	21	AA622377
11	279.8	11.9	2248	22	AA503053
12	276.8	11.8	3525	23	AA509775
13	177.8	7.6	3319	15	AA070152
14	177.8	7.6	3319	17	AA742166
15	164.4	7.0	2034	23	ABL16733
16	164.4	7.0	4034	23	ABL16732
17	161	6.8	15532	23	ABL12290
18	134.8	5.7	12075	23	ABL14182
19	124.4	5.3	758	22	AA123900
20	124	5.3	1407	22	ABA17294
21	118.6	5.0	459	21	AA602955
22	114.8	4.9	2111	21	AA599916
23	112	4.8	5884	22	AA545037
24	110.6	4.7	5368	22	ABA83095
25	110.6	4.7	5368	22	AA697834
26	110.4	4.7	8848	23	ABL09774
27	107.8	4.6	3572	19	AA67162
28	106.4	4.5	1230	24	AA623972
29	106.4	4.5	4757	24	AA623970
30	106.4	4.5	4790	22	AAH72653
31	106.2	4.5	1950	22	AAH15830
32	106.2	4.5	3720	23	ABL13337
33	104.6	4.4	814	22	AAH05392
34	104	4.4	9626	20	AA526819
35	103	4.4	1053	19	AAV47580
36	103	4.4	1053	24	AA566044
37	102.6	4.4	5607	22	AA545225
38	100.4	4.3	8083	22	AA570692
39	100.4	4.3	8503	22	AA158253
40	99.6	4.2	520	22	AA181860
41	99	4.2	4550	23	AA584836
42	98.2	4.2	1391	22	AAH89871
43	97.2	4.1	4557	21	AA646311
44	97.2	4.1	4989	21	AA642969
45	95	4.0	1128	24	AA623971

#### ALIGNMENTS

RESULT 1

AAx87656 standard: DNA: 2352 BP.

AAx87656;

26-OCT-1999 (first entry)

Thermomyces lanuginosus kinesis motor protein TL-gamma DNA.

TL-gamma: kinesis; motor protein; microtubule; unc-104; infection;

neurodegenerative disease; Alzheimer's disease;

Parkinson's disease; Huntington's disease;

amyotrophic lateral sclerosis; ss.

Thermomyces lanuginosus.

WO9937659-A1.

29-JUL-1999.

22-JAN-1999; 99WO-US01355.

23-JAN-1998; 98US-0072361.

(REGC ) UNIV CALIFORNIA.

Goldstein LSB, Sakowicz R;

WPI: 1999-493950/41.

P-PSDB: AAY06618.

Human secreted pro

Human diagnostic a

Drosophila melanog

k39 gene of Leishm

Leishmania chagasi

Drosophila melanog

Drosophila melanog

Drosophila melanog

Drosophila melanog

Drosophila melanog

Human breast cancer

Human nervous syst

Human secreted pro

Polynucleotide enc

cDNA encoding nove

KIA0591 ovarian t

Human neuroblastom

Drosophila melanog

Drosophila kinesis

Human HsK1f5 prot

Human kinesis supe

Human cervical can

Human cDNA sequenc

Drosophila melanog

Human cDNA clone (

Nucleotide sequenc

Leishmania antigen

Leishmania antigen

cDNA encoding nove

DNA encoding novel

Human polynucleoti

Human polynucleoti

DNA encoding novel

Human bone marrow

Arabidopsis thalia

Arabidopsis thalia

Human HsK1f5 prot

PT New nucleic acid encoding microtubule motor protein, used for  
PT diagnosis of fungal infection and neurodegenerative disease  
XX  
PS Claim 4: Page 69-70; 75pp: English.

CC This is the DNA sequence of the Thermomyces lanuginosus gene  
CC coding for TL-gamma (see AY0618), a novel ATP-dependent, plus  
CC end-directed microtubule motor protein that is a member of the  
CC unc-104 family and kinesin superfamily. The gene was isolated  
CC from genomic DNA by PCR amplification (see also AX87659-61).  
CC The invention provides TL-gamma nucleic acids, proteins and  
CC antibodies, and methods of screening for TL-gamma modulators  
CC potentially useful for treating hyphal fungal infections and  
CC diseases caused by mutated TL-gamma, e.g. neurodegeneration  
CC involving anterograde axonal transport, such as Alzheimer's,  
CC Parkinson's or Huntington's diseases or amyotrophic lateral  
CC sclerosis. TL-gamma nucleic acids are also used to identify  
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or  
CC other motor proteins, to generate protein structural models, for  
CC recombinant production of TL-gamma, as antisense molecules, to  
CC produce transgenic or knockout animals, and in gene therapy.  
CC Detection of TL-gamma allows differentiation between hyphal and  
CC non-hyphal fungal infections.

SQ Sequence 2352 BP; 626 A; 564 C; 672 G; 490 T; 0 other;

Query Match 100.0%; Score 2352; DB 20; Length 2352;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcggcggttgaaatatacaagtggtgtgctgggttacgcccgttcaacgcccgaaga 60  
DB 1 atgtcggcggttgaaatatacaagtggtgtgctgggttacgcccgttcaacgcccgaaga 60  
QY 61 atgacccgttgcgcaaaatgtattgtgcgatatggaagaatacaaccatccctc 120  
DB 61 atgacccgttgcgcaaaatgtattgtgcgatatggaagaatacaaccatccctc 120  
QY 121 cctcgggttcgcaagaagaagcgcgtaaagtgtgcaaaactattatggatggccgaag 180  
DB 121 cctcgggttgcgcaagaagaagcgcgtaaagtgtgcaaaactattatggatggccgaag 180  
QY 181 gcatctgcttcgcatcggtcgtatgtgtcctttgacaaagaatgctcccaactatgcaga 240  
DB 181 gcatctgcttcgcatcggtcgtatgtgtcctttgacaaagaatgctcccaactatgcaga 240  
QY 241 caaggaagactattccaagatcgcgagtcgccgtcttgataatgatccaaggttat 300  
DB 241 caaggaagactattccaagatcgcgagtcgccgtcttgataatgatccaaggttat 300  
QY 301 aacaatgtatcttcgctacggttcagacgcggttcggcgcaagtccctatcaatgatggc 360  
DB 301 aacaatgtatcttcgctacggttcagacgcggttcggcgcaagtccctatcaatgatggc 360  
QY 361 tatggcaagaagacatgctgtatccgcggaatttgcgaagacatggttcggcggtat 420  
DB 361 tatggcaagaagacatgctgtatccgcggaatttgcgaagacatggttcggcggtat 420  
QY 421 gaacttcgagaagaagaacacccacttcgacgcgttcgaagtttcgttcttggaaattac 480  
DB 421 gaacttcgagaagaagaacacccacttcgacgcgttcgaagtttcgttcttggaaattac 480  
QY 481 aatgaacgaatgtcgagaactgtgtaatcgttcgacaagaagggaatcgaagtccgagaa 540  
DB 481 aatgaacgaatgtcgagaactgtgtaatcgttcgacaagaagggaatcgaagtccgagaa 540  
QY 541 caccgcgtcgacgcgcccctacgttggagacattggcggaagctgtgctgtgcatcattcaa 600  
DB 541 caccgcgtcgacgcgcccctacgttggagacattggcggaagctgtgctgtgcatcattcaa 600  
QY 601 gaatcgaaatctcatgtagtgagggcaacaagccgaagcgttgcgcgcacaaatg 660  
DB 601 gaatcgaaatctcatgtagtgagggcaacaagccgaagcgttgcgcgcacaaatg 660

DB 601 gaatcgaaatctcatgtagtgagggcaacaagccgaagcgttgcgcgcacaaatg 660  
QY 661 aacgagacatccagtcgatacccaagcgtcttcaacttgaacttgaagcaaaagtggcat 720  
DB 661 aacgagacatccagtcgatacccaagcgtcttcaacttgaacttgaagcaaaagtggcat 720  
QY 721 gatgaagaagccaaatgtgacacagagaaggtttgcgaagatcagctgtgatatgttcg 780  
DB 721 gatgaagaagccaaatgtgacacagagaaggtttgcgaagatcagctgtgatatgttcg 780  
QY 781 ggttctgagcgagcaagctccacccggaagctacgtgagcgcgactggaagggtgtcaga 840  
DB 781 ggttctgagcgagcaagctccacccggaagctacgtgagcgcgactggaagggtgtcaga 840  
QY 841 atcaacgcgcactttcgaccctcctgctgtgtatttcagcgcgcagggatatgtctcg 900  
DB 841 atcaacgcgcactttcgaccctcctgctgtgtatttcagcgcgcagggatatgtctcg 900  
QY 901 ggaataacagaagaagaatcagttagtaacttaccgagatccggttactgacgtgtctctg 960  
DB 901 ggaataacagaagaagaatcagttagtaacttaccgagatccggttactgacgtgtctctg 960  
QY 961 aaggaactccttggaggagcaactcgatgacccgcatgatttgcgcacatttcgctgtgat 1020  
DB 961 aaggaactccttggaggagcaactcgatgacccgcatgatttgcgcacatttcgctgtgat 1020  
QY 1021 attaactttgaagagactctcagtaacctcctcgatatcgagacttcggaagcgatcaag 1080  
DB 1021 attaactttgaagagactctcagtaacctcctcgatatcgagacttcggaagcgatcaag 1080  
QY 1081 aacacgcagtggttcaatgaagaacccgacgcggaatgacgcggaagtgtgaaggagaa 1140  
DB 1081 aacacgcagtggttcaatgaagaacccgacgcggaatgacgcggaagtgtgaaggagaa 1140  
QY 1141 ctgcgcgagcttgaggagcaaatcccaagcagtggtgagggtggaagtggtgcagagagt 1200  
DB 1141 ctgcgcgagcttgaggagcaaatcccaagcagtggtgagggtggaagtggtgcagagagt 1200  
QY 1201 tctgcgcgccagtggtgaggaatcgtacccgccacacgcgcgcgcgcgcgcgcgcgcgcgc 1260  
DB 1201 tctgcgcgccagtggtgaggaatcgtacccgccacacgcgcgcgcgcgcgcgcgcgcgcgc 1260  
QY 1261 tcgattcaagcagccggaatgcgacagtcgaagaanaatgaagaagcagaatcgttgagcaaa 1320  
DB 1261 tcgattcaagcagccggaatgcgacagtcgaagaanaatgaagaagcagaatcgttgagcaaa 1320  
QY 1321 ctggaacagagtggaagcgtctacgtgatactcaataagacccctgggaaggaagctggcc 1380  
DB 1321 ctggaacagagtggaagcgtctacgtgatactcaataagacccctgggaaggaagctggcc 1380  
QY 1381 aagcgcgaggaatccacaagaagcgaagcggcgctcgagaggtcgggtatcagcatc 1440  
DB 1381 aagcgcgaggaatccacaagaagcgaagcggcgctcgagaggtcgggtatcagcatc 1440  
QY 1441 gaaaagggtctgttgcgccttaccactccaagaanaatgccaactatgactatgagc 1500  
DB 1441 gaaaagggtctgttgcgccttaccactccaagaanaatgccaactatgactatgagc 1500  
QY 1501 gatgatcctcttctgctgt 1560  
DB 1501 gatgatcctcttctgctgt 1560  
QY 1561 ggaacgcgtcaacccaagatacaagaagcggaaattcgttgaaaggttcgaaatccttgaaa 1620  
DB 1561 ggaacgcgtcaacccaagatacaagaagcggaaattcgttgaaaggttcgaaatccttgaaa 1620  
QY 1621 gaacactgtacglttgaaaatgttgacaacglttgaccatcgtgccaacagagaagct 1680  
DB 1621 gaacactgtacglttgaaaatgttgacaacglttgaccatcgtgccaacagagaagct 1680  
QY 1681 gctgtcatgttgaaagcgtggtggaatcgacaagaagccttcgctccgacggtctacagg 1740  
DB 1681 gctgtcatgttgaaagcgtggtggaatcgacaagaagccttcgctccgacggtctacagg 1740

Oy	1741	atctctctggcgatcttccacattttcgaattcaacccatccggaaagagtcgtcggaa	1800
Db	1741	atctctctggcgatcttccacattttcgaattcaacccatccggaaagagtcgtcggaa	1800
Oy	1801	cggaagaagaataatccttgccttcgcacatttcgtcaccacaagtcagttgggttcgcgtc	1860
Db	1801	cggaagaagaataatccttgccttcgcacatttcgtcaccacaagtcagttgggttcgcgtc	1860
Oy	1861	ccaggccgttcacgacccggacactggaacagcggtctcgatctgcgacgctgattctcgc	1920
Db	1861	ccaggccgttcacgacccggacactggaacagcggtctcgatctgcgacgctgattctcgc	1920
Oy	1921	tcagatctctccttccgcgcacttcgttggaagaagatagcagctagttctctatgtccgag	1980
Db	1921	tcagatctctccttccgcgcacttcgttggaagaagatagcagctagttctctatgtccgag	1980
Oy	1981	gaagctgtctagcgcgcattccctaaggttggaattcagaagaagctctcatcttcacagatgcag	2040
Db	1981	gaagctgtctagcgcgcattccctaaggttggaattcagaagaagctctcatcttcacagatgcag	2040
Oy	2041	ttggatgcatatttcgaagatgttcgaagaagcgcggtcagttcgtctggtcgtcga	2100
Db	2041	ttggatgcatatttcgaagatgttcgaagaagcgcggtcagttcgtctggtcgtcga	2100
Oy	2101	gacacaagaagatgcgattctgcgaagattcgtttccggtctcgttgacaaatacatgtccaat	2160
Db	2101	gacacaagaagatgcgattctgcgaagattcgtttccggtctcgttgacaaatacatgtccaat	2160
Oy	2161	ggaacacttgataattcttcgcgcgcgtacacgcacattactatgcgggtacacccctcgagt	2220
Db	2161	ggaacacttgataattcttcgcgcgcgtacacgcacattactatgcgggtacacccctcgagt	2220
Oy	2221	gatgaacgacggtgacgcgctgtttttttgtgtgaataagaagtcgaaacagatgcgtcaat	2280
Db	2221	gatgaacgacggtgacgcgctgtttttttgtgtgaataagaagtcgaaacagatgcgtcaat	2280
Oy	2281	gtttatcttgaagaagtttgcgtcaacagcggtctcagatggaaagaagccctgaaaacgcg	2340
Db	2281	gtttatcttgaagaagtttgcgtcaacagcggtctcagatggaaagaagccctgaaaacgcg	2340
Oy	2341	aagcagaagaatc 2352	
Db	2341	aagcagaagaatc 2352	
RESULT 2			
ABL4183	ID ABL14183 standard; cDNA; 5322 BP.		
XX	AC	ABL4183;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 37031.	
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ss.		
XX	OS	Drosophila melanogaster.	
XX	PN	WO2001/1042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-DS09231.	
XX	PR	23-MAR-2000; 2000US-191637P.	
XX	PR	11-JUL-2000; 2000US-0614150.	
XX	PA	(PEKE ) PE CORP NY.	
XX	Venter JC, Adams M, Li PMD, Myers EM;		
PI			

[illegible]

QY	674	gtctgattcccaacgcgcgtcttccactcttgtaaccttggaagcgaaaggtgcatgatgaaagagcca	733
Db	731	ccgcgtctctacgacgcgtctcttaacaactctctccaccaacgctgcgtacatgctgatgaa	790
QY	734	aaatgtagcaacaagaagaggttgcgaaagatcagtcgctgtagatatttggcgggtcttcgacgga	793
Db	791	atttgaccacgaanaaggtatcccaagataagcttggttagacttggccgggttcgaaacgag	850
QY	794	caacgctccacggagactactgtagcgagcgacatcgaaagggaggtgcagagatcaacgcgtcac	853
Db	851	cgagattccacgtgagccaaaggacactgcgtctgaaagggagggagcgacaatatacaatctct	910
QY	854	tttcgacccctaaagtctgtagttcgcaacgcgtcagctagagtagatctggtcgggaaacag----	909
Db	911	tgcaccatctgggcaaagtattctccagcttggttgagggaagtggtcttccaaagaaagaagaa	970
QY	910	--aagaagaatagttagttacaccttaacgagatctcgtagtacggtggctctgaaagact	967
Db	971	ccaaagagagcagattttatctccgtacacgttgattctgcgtctgaaacctgctgttagcgaa	1033
QY	968	cccttgagaggaacactcgaatgacccgcacatgatttgcgcacatttcgcgtctgatttaaat	1027
Db	1031	acttggagaggaattccgaaagacagtatgatttggtgcgtacttcaacgaacagatataaact	1091
QY	1028	ttgaaagagactctcagtacaccttcgatatactgcgagactctcgcgaaagcgatataaagaacag	1087
Db	1091	acgaatgaaacccctccagcacacatgcgtactatgcgtagctgctgcacaagaaattggttagaag	1150
QY	1088	caagtgtgcaatgaaaaccccgaaacgcgcgagatgcacccgcgggttgaaagggagaaatcgcgc	1144
Db	1151	ctggtgtgcaatgaaagatgacatgacatgcagactgcgtattcgcgaaactcaagaggagagatccaa	1210
QY	1148	agactagagagcaaacctccagagcaagtggtggaaggttgaaaggttggtgcagagaggtcttcgag	1207
Db	1211	agctctcgagacttactcacaagccgagagcattgaaagtgcagagaggaacccgattgcaagag	1270
QY	1208	ggccagtggaaggaatcgtgacccgcgcgacacgcgcgtccgaaaggaagcaaatcgtgtcgattc	1267
Db	1271	tgtgtgtggaagagcgagatgcygataaagaaatgagctccaccaaaagctcaacggttgatataagt	1330
QY	1266	agcagccgagatgc--gaaacagttcaagaaatgagcaagagcaagaaatcgtgcgagcaacttga	1324
Db	1331	cgccacataagttcacgtatcgtataatgtagtcccaacagcgagatgagtggtgtgagtcacgtgc	1390
QY	1325	accagagatgaaagagctctatctatctgagatctccatcaagactcgggaagagaaactcgtccga	1384
Db	1391	agggccagcgaaactctatctatgcaagctccaaagacactcgtgagggagaaacttaagcgca	1450
QY	1385	ccgagagaaatcccaaaagaaagaaagacgcgtctcgagagagctcgggttatcagatccgaa	1444
Db	1451	ccgagagagattcgcgtgcagagagagcggtctctgcgaaataggggtggtgtgtctcaag	1510
QY	1445	agg-----gcttgttggtgccttaccacttccaaagaaatgccaatctagtacaaacttga	1498
Db	1511	aagatggaataacagcttgcgtatcttccacccgaaagaagatccgcacatttgcgttccaaactaa	1570
QY	1499	gcgatgacccctcttcgctgcgtatgagtgcttctgcataacaataacgcccgcgagcaaaag	1558
Db	1571	acgagagatcccaactctgtcgtgagtgctgttactaacatcaagaaggggtccaaactcgt	1630
QY	1559	ttggaagaactcaaccaagatacaacaagcggaaatctcgtctgaacggttctgaaagactctga	1618
Db	1631	tgggttaccacatgaaagcaaaatgctgcccaagacatacgtcttccgatctgcacatccca	1690
QY	1619	aaggaacatgtagctttagaaatgtgagacaagcttgtgaccatctggtcccaacgagaa	1678
Db	1631	aagagcaactgtacactttagaagaacaagaacagcgcgttaccctgcgtgcacacaagaagt	1750
QY	1679	ctgctgtcatgtgtgaacgggtgctgaaatcgacaacagctactcgcctccgcagcggctaca	1738
Db	1751	ctatcatctatgtaaatgagcgcaaggtgtgtgaaccggaaagttcttaagaccggtcttc	1810
QY	1739	ggatcatctcggcgagatttccacaatttccgatccaacatccggaagaaagctcgttgcg	1798

















CC in disease diagnosis and population studies.

XX in disease diagnosis and prevention studies.  
XX  
SQ Sequence 10682 BP; 2817 A; 2443 C; 2602 G; 2812 T; 8 other;

Query Match 13.4%; Score 316.2; DB 21; Length 1082;  
Best Local Similarity 59.6%; Pred. NO. 8.9e-84;  
Matches 616; Conservative 1; Mismatches 389; Indels 27; Gaps 4;

Qy	157	aaaactatgatgagccggaagcatttgcgttcgactcggctcgattggctccttgac	216
Db	307	aaagatccaaaggagctccaaagtccttcagcttcgactattcctactggtctcatacc	366
Qy	217	---aagaatgctcccaactatcgagacaggaagaccctattccaaagatcctcgagtcctcg	273
Db	367	tcaccggaagatccccttgcttcgactcctaaaccgctggtacaaatggcaagaa	426
Qy	274	cttcgtgataatgattccaaggttatacaaatgtatcttcgctacoggtcgacacoggt	333
Db	427	atgctcttacacgcttggaggatataatgctgtatttttgcstatggcgagactggt	486
Qy	334	tcggggcaagtctctattcaatgatgggctatggcaagg-----agcatggcggtgatcccg	387
Db	487	gctggaaaatcttatcaatgattgggttaaacaggaagaaagccaggtcggtcatctcca	546
Qy	388	cgaatttgcagacatgttcggcggtattaatgaactgcagaaggacaagaacctcaat	447
Db	547	cagttaagtgaagactctttgagaaaaatcaatgacaactgtaatgaagaaa---tgtct	603
Qy	448	tgcacgcgtgaagtcttcgtacttgaaatttacaatgaacgagtcgcgagacttgcgaat	507
Db	604	tactctgtagagtgagctacatggaaatttactgtgaagagtcagagatttgtgaat	663
Qy	508	ccgtcgacaaaggggaatctcaaggttcgagacaccccgctcgacggccctcagctggag	567
Db	664	ccaaaacaaggttaatttgctgctgagcaccacactcttgaccctatgtggag	723
Qy	568	gacttgcggaagctgctgctgagatcattccaaagaaatcgaaaatctctatggatggggc	627
Db	724	gattctccaagttggcagttacttctcacagacattgctgacctcaatggatgctggg	783
Qy	628	aacaaagccagaaacggttccgcccacaaacatgaacgagacatccagtcgactccacgcc	687
Db	784	aacaaagccagagacgtggcagctacaacatgaaatgaacaaagtgcgctccacgct	843
Qy	688	gtcttcaactttgaccttgacgcaaaagtggcagtgaagagagacccaaaaatggacacagag	747
Db	844	gtgtttacgatttttcccccagaagaacacgataatgagacacacttccacttcacgtag	903
Qy	748	aaggttcggaagatcagctgttagatttggcggttctgagcgagcaacgtccacccgga	807
Db	904	aaggtcagtaaaatcagcttggctggatctagcaggaaagtgaacgagctgattcaacttgg	963
Qy	808	gctactggagcgcgactgaagaggggtcgagagatcaaccctcactctcgacccctaggt	867
Db	964	gccaaagggactcgattaaaggagagcaaatattataatgaagtctttacaactttggc	1023
Qy	868	cgctgattgcagcgtacgcga-----tatgtcgtcgggggaaaaacagaag	912
Db	1024	aaagtcatttcagccttggccgaggtgggaactgcactagcaagagtaaaagaagaag	1083
Qy	913	aagaatcagttagtactcacgagattcgttactcagctggtcttcgagagactccttg	972
Db	1084	aaaacagattttatccctacagggattctgacttacttggctctcgagaaaaattta	1143
Qy	973	ggaggaactcgtatgaccgcctatgttcgcgcattctgcgctgctgatatcaactttgaa	1032
Db	1144	ggtggcaattctcgactgcaatggttctgctctgagccccgcggatatcaactacgat	1203
Qy	1033	gagactcagttacccttcgatatacgggactctgcgagcgaaatcaagaacacgcagtg	1092
Db	1204	gagactttgagcactctgagatgatacgatcgtgcgaacaaaattaaatgaaatgctgtt	1263

Qy	1093	gtcaatgaagaccgaacgcgcgcgagttgaaggagggaactcgcgcagctg	1152
Db	1264	atcaatgaggaccaccaatgccaaactggtctgtaattaaaggaggaggtgacacgcgtg	1323
Qy	1153	aggagcaaacctc	1165
Db	1324	aaggacctcttc	1336
RESULT 10			
AAF22377			
ID	AAF22377 standard; cDNA; 1701 BP.		
XX			
AC	AAF22377;		
XX			
DT	26-MAR-2001 (first entry)		
XX			
DE	Human secreted protein gene 5 SEQ ID NO:15.		
XX			
KW	Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;		
KW	antirheumatic; antiproliferative; cytostatic; cardant; vasotropic;		
KW	cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide		
KW	fungicide; ophthalmological; gene therapy; autoimmune disease; neoplas		
KW	rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;		
KW	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;		
KW	angiogenesis; nervous system disorder; Alzheimer's disease; infection;		
KW	ocular disorder; corneal infection; wound healing; skin aging;		
KW	food additive; preservative; ss.		

RESULT 10  
AAAF22377  
ID AAF22377 standard; cDNA; 1701 BP.  
XX  
XX  
AC AAF22377;  
XX  
XX  
DT 26-MAR-2001 (first entry)  
XX  
DE Human secreted protein gene 5 SEQ ID NO:15

Human; secreted protein; diagnosis; immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm; rheumatoid arthritis; hyperproliferative disorder; cardiac arrest; cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia; angiogenesis; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection; wound healing; skin aging; food additive; preservative; ss.

OS Homo sapiens.

XX PN WO200061629-A1.

XX  
PD  
19-OCT-2000XX  
DE 06-APR-2000-

XX  
PP 00-APB-1000: 0015-013000

PR 20-JAN-2000; 2000US-0176931.

PA (HUMA-) HUMAN GENOME SCI INC

XX

XX

DR P-PSDB; AAB63138.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition

PS Claim 1: Page 419-420: 533pp: English:

AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134 to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins and polypeptides homologous to them. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; and ophthalmological. The polynucleotides and proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for







PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions -

XX

PS Claim 1: SEQ ID NO 23807: 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](ftp.wipo.int/pub/published_pct_sequences).

XX  
SQ Sequence 3525 BP; 896 A; 914 C; 1009 G; 706 T; 0 other;

Query Match 11.8%; Score 276.8; DB 23; Length 3525;  
Best Local Similarity 57.7%; Seed No. 348-73

best local similarity 37.7%; Fred. NO. 3.4E-12;  
Matches 576; Conservative 0; Mismatches 407; Indels 15; Gaps

QY 179 aggcatttcgattcgaatcggtcgtatctggtcctttgacacgaatgctcccaactatcga 238

98 acgacttcacctcgactactccctatttggtcattcgatgcgcaggaatccgcacttcacca 157

[illegible]

QY 239 gacaggaaagacccatccccaaagacctcggagtcctccggataatgcattcaagggctc 298

Db 158 ccaggagcagggtttacagcgatc199ggaatgatgtcgtggattgcgcctacgaaggat 217

Qy 299 ataacaattgtatatcttcgcctacggtcagaccggttcggaagtcctattccaatgatgg 358

Query Match 11.88; Score 276.8; DB 23; Length 3525;  
Best Local Similarity 57.7%; Pred. No. 3.4e-72;  
Matches 576; Conservative 0; Mismatches 407; Indels 15; Gaps 4;

179 aggcatttgccttcgattcggctcgtatttgcctccttctgacgaagaaatcctcccaaatatcggg 338

**Nv** 179 aggcattttgacnttcgatcagatccttatctgtccttcttaaacaaagaaactacctccccactataccga 338

27 2/3 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

100

Db 98 acgacttcaccttcgactactcctatttggtcattcgatcgaggatccgcacttcgcca 157

Qy 239 gacaggaagacctattccaagatctcgaggtccgcttctggataatgcattcaagggtt 298

Dbb 158 cccaggagcaggtttacagcgaatctagcaaatgattcattgattacacctacgaaggat 217

-----

[illegible]

**У**

Db 218 acaatgcgtgtgttttgcctacggacagactggctctggcaagaccttcaccatgatgg 277

Qy 359 gctatgqcaaggagcatgqccgtatccqccqatattccagqacatattccqccatatta 418

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

Dh 278 acacgccaacaatccgggcctaataccgggcacatctgcgaaggacatctttaccccaatgc 337

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

110

419 atgaaactgcagaaggaacgaagaccctcacttgcaccgctcgaagttctcgtaacttggaaattt 478

[illegible]

Db 338 gtg --- tgggccaggagtcgggcactgggtacaggacgcacgtccagctatctgggagatct 394

Ov 479 acaatgaacgaatgccgaacttactgaatccatccacaaaggggaat---ctcaaggtcc 535

[illegible]





Db	865					-----gga	aaagc	ccccata	906
Qy	928	ccttaccgagattcggtactgacgtggcttctgaaggactccttgaggaggaactcgcgtg	987						
Db	907	ccgtatcggaactccaagttaactcgacttctgcaggactcctcgggtgtaactccaag	966						
Qy	988	accgccatgattgcgcacatttcgcctgctgatattaaactttgaagagactctcagttacc	1047						
Db	967	accgtcatgtgtgccaccatttagtccagcggatagcaattatattggaactatataccaca	1026						
Qy	1048	cttcgatatcgggactctgcgaagcgaatcaagaaccacgcagtggtcaatgaagaccg	1107						
Db	1027	cttcgctatgctagtcgtgctaagaataccaataatcgcatgcacatcaatgaggaaccc	1086						
Qy	1108	aacgcgcggatgatccgcgagttgaaggaggaactcgcgcagctgagagcaaaactccag	1167						
Db	1087	aaggatgcttcttcgtcgtcactttcaggagagattgccgcacttcgcgaagcaattggag	1146						

Search completed: July 25, 2002, 05:44:40  
Job time: 6948 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	380.6	16.2	3930	3	US-09-162-373-2	Sequence 2, Appli
2	380.6	16.2	3930	4	US-09-467-946-2	Sequence 2, Appli
3	283	12.0	4176	4	US-09-722-139-1	Sequence 1, Appli
4	284	11.2	1077	4	US-09-722-139-3	Sequence 3, Appli
5	177.8	7.6	3319	1	US-08-006-6768-2	Sequence 2, Appli
6	177.8	7.6	3319	1	US-08-282-845-1	Sequence 1, Appli
7	177.8	7.6	3319	2	US-08-428-414A-4	Sequence 4, Appli
8	177.8	7.6	3319	5	PCF-US94-00334-2	Sequence 2, Appli
9	177.4	7.5	1689	4	US-09-718-841-1	Sequence 1, Appli
10	158.2	6.7	1070	4	US-09-718-841-3	Sequence 3, Appli
11	118.2	5.0	29793	4	US-09-302-812-38	Sequence 38, Appl
12	118.2	5.0	29793	4	US-09-511-477-38	Sequence 38, Appl
13	107.8	4.6	3572	2	US-08-713-815A-2	Sequence 2, Appli
14	106.4	4.5	1230	4	US-09-572-191-5	Sequence 5, Appli
15	106.4	4.5	4757	4	US-09-572-191-1	Sequence 1, Appli
16	95	4.0	1128	4	US-09-572-191-3	Sequence 3, Appli
17	80	3.4	3709	4	US-09-541-782-7	Sequence 7, Appli
18	79.4	3.4	3883	1	US-08-468-036-33	Sequence 33, Appl
19	79.4	3.4	3883	2	US-08-376-843-33	Sequence 33, Appl
20	79.4	3.4	3884	4	US-09-541-782-3	Sequence 3, Appli
21	78.4	3.3	2389	1	US-08-480-552-18	Sequence 18, Appl
22	78.4	3.3	2389	2	US-08-486-382-4	Sequence 4, Appli
23	78.4	3.3	2389	3	US-09-235-546-4	Sequence 4, Appli
24	78.4	3.3	2389	3	US-08-929-208-18	Sequence 18, Appl
25	78.4	3.3	2389	4	US-09-568-315-18	Sequence 18, Appl
26	71.2	3.0	3741	4	US-09-541-782-9	Sequence 9, Appli
27	70.6	3.0	897	4	US-09-621-233-1	Sequence 1, Appli

```
Db 347 cgtgtgcatctttgctctatgggcagacccgggctgggaaatcctataccatgatggggcg 406
QY 360 ---ctatggcaagagcatgctgtgatcccgcggtattgccaggacatgttccggcgctat 416
Db 407 acaggagccaggcgagcagggcatcgtgcccgactctgtgaggaccttcttctcgcgt 466
QY 417 taatgaactcagaagagacaaagacacctcaattgcacgcgtcgaaagtcttcgtacttggaaat 476
Db 467 tagtgaacacagagtctcag---ctatcctactctgtgaggtgagctataggat 523
QY 477 ttacaatgaacagagtcgagacttctgaaatccgtcgacaaaggggaatctcaaggtccg 536
Db 524 ctactgtagcgggtgagacacctctgaaccccaagtcgggttctctcgggcccg 583
QY 537 agaacaccctcgcacccgcccctacgtggaggacttggaagctgtgctgcatatt 596
Db 584 ggagcaccctatcctggcccgctacgtgcagacctgtccaaattgctgtgacctcta 643
QY 597 ccagaataatcgaaatctcatgatgagggcacaacaaagccagaaacggttgcgcgcacaa 656
Db 644 cgcagacattgctgacctatgactgtggaataaaagcagcagactgtggtgcaccaa 703
QY 657 catgaacgagacatccagtcgactccacgcgcgtcttcactttgaccttgacgcgcaaaagt 716
Db 704 catgaatgagaccagcagcgttcccatcgcgtctttaccatcgtcttcacacagcgtg 763
QY 717 gcatgataagagacccaaaatgacacagagaaggttgcgaagatcagttcgtgtagatt 776
Db 764 ccatgacagctcacgggctggactcggagaaggtcagtaagatcagttgttggacct 823
QY 777 ggcggttctgagcagcaacgtccacggagctactgtgagcgcgactgaagagggtgc 836
Db 824 tgcgtggagtgaagcagccgactcctcaggggcggccgggcatgggctgaaggaagagc 883
QY 837 agagataccgcgtcactcttcgcaccttaggtcgtgtattgacgctagcgtagcatatgc 896
Db 884 caacataataagctcgtgactacactagggaaagtgtctcgtcccttcagatatgca 943
QY 897 gtcgggaaaaacagaagaatcagttagttaccttaccgagattcgtactgactgtgct 956
Db 944 atc---aaagaagcgaaagtcggtatttatccctacacagggactctgtcactggt 1000
QY 957 tcgaaagactcttggaagcaactgatgacgcgcactatgtgcgccttttcgcctgc 1016
Db 1001 gctcaaggaaatttggggggaactcacgcacagcagcatgctgacccctgagccctgc 1060
QY 1017 tgattataactttgaagagactcagtaaccttcgatatgcgacttcggaactctcgaaagcaat 1076
Db 1061 tgacataattacgagagactctcagaccctcaggtatgctgacgcgacccaagcaaat 1120
QY 1077 caagaaccacgcagtggttcaatgaagaccgcgaacgcgcgagatcgcgcgagttgaagga 1136
Db 1121 ccgctgcaatgccatcatcacgaggaacctaatgcccgcgtgattagagagctgcagga 1180
QY 1137 ggaactcgcagctagagagcaaat-----ccagagcagtggt 1176
Db 1181 ggaagtgcgcgcgtcgggaactgctgattgctcaggggactgtcagcctctcgtctgga 1240
QY 1177 ggaagtgaggtgtgagaggttct-----ggcgggccagtgaggaatcgtacc 1229
Db 1241 aggcctgaagacagaaagagaggtgtcagagggcgcctcgcagctgtgtcatctcccc 1300
QY 1230 gccgcacacgcctcgtcagagaacaaatcgtgtcgtattcagaccgagctgcagactc-- 1287
Db 1301 agctccagtttccacctcatccaccacacataatggggagctgagcgtcattctc 1360
QY 1288 -----aagaaatgagcaagcgagaaatcgtggagcaactgacaccagatga 1334
Db 1361 ccccaacacggagtccacagattgggctgaggagccatggagagctgcagagagacaga 1420
QY 1335 gaagctctatcgggtctcaatcagactcgtggagagaagctggcgaagaccgaggaat 1394
Db 1421 gaagattatagctgagctgaacgagacatgggaggaagactcagcgaagacagaagccct 1480
```

```
QY 1395 ccacaagaaacgagaagcggcgctcgaggagctgggtatcagcatc-----gaaaagg 1448
Db 1481 gaggatgagagagaagcattgctgctgagatggggtggtccggtccggaggatggggg 1540
QY 1449 ctttgtggcccttacacatcccaaaagaaatgccacatctagtcaacttgagcgtatcc 1508
Db 1541 aactgtggcgctctctccaaagaagactcccaactggtgaacctgaacgaagaccc 1600
QY 1509 tcttctggctgagtgcttctctacaacatcaagcccgccagacaaggttgaaacgt 1568
Db 1601 tctgactcgtgagtgctctctaccatcaaaagatggcgctcaccagggtcggcccaagt 1660
QY 1569 caacaagatacacaaagcggaaattcctgaacggttccgaagatc---ctgaagaaca 1625
Db 1661 agatatggacatcaagctgaccggacagtcttcggagcaacacactgtctgttccggag 1720
QY 1626 ctgtactgttgaaatgtggacaacgttgtgacctcgtgccaaacgagaaggtctgt 1685
Db 1721 catcccccagcagatgagaagtggtgctcactctgagcctgtgaaagagctgagac 1780
QY 1686 catggtgaacggcgtgcgaatcgacaagcactcctcgcgcagcgcgtacagatcat 1745
Db 1781 atatgtaatgggaagcttgtgacgggagcgcgtgtgtgtaagtcaggggaataggattgt 1840
QY 1746 cctgggggatttccacatttttcgattcaaccatcccgaaagaagctcgtcggaacgca 1805
Db 1841 gatgggcaagaacacagctttccgcttcaaccacccggagagcgcaaggtcggaaacgga 1900
QY 1806 a 1806
Db 1901 a 1901

RESULT 2
US-09-467-946-2
; Sequence 2, Application US/09467946
; Patent No. 6248594
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/467,946
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/162,373
; EARLIER FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1281811
US-09-467-946-2
```

Query Match 16.2%; Score 380.6; DB 4; Length 3930;  
Best Local Similarity 55.0%; Pred. No. 2.6e-103;  
Matches 1023; Conservative 0; Mismatches 739; Indels 99; Gaps 10;

```
QY 6 gggcggtggaaatatcaagtggtggtcggtgtacgccgttcaaccccgagaaatcga 65
Db 80 ggcgtggtcctcgtgaaagtggcagtgggttcggcccttcaaccccggtgagccag 139
QY 66 ccgtggcgcaaaatattgtgcggatggaaggaaatcaaacatctctccctcc 125
Db 140 ccagatgccaagtgtgtgtgctcagcatgcagggcaacacccctccatcatcct-- 197
QY 126 gggtcggaagagaagcgcgtgtaaaagtgcgaactattatggtgcccgaagcatt 185
```



Db 11 tcaagggtgcgtgaggttcgcccgaatgaatcgagggaaagacttgagggccaagt 70  
QY 80 gtattgtcgcatggaagaaatacaaccatctccaccctccctccgggtgcggaagaga 139  
Db 71 tcattattcagatggagaagaacaaacgacaaatcaaaacttaaaatataccagaag--- 127  
QY 140 aggcgcgtaaagtggcaaaactattatggatggcccggaaggcatttgcgttcgcatcgg 199  
Db 128 -----gaggcacttgggactcaggaagagacggaccaagaccttcacatgactttt 181  
QY 200 cgtattggctttgacaagaatgctcccaactatgagacagagaagacactattccaag 259  
Db 182 cttttattctgctgatacaaaaagccagatcactcggttcacaagaagaatggtttccaaa 241  
QY 260 atctcggagtcgcccttctgataatcattcaagggttataaacaattgtatcttcgct 319  
Db 242 ccttcggcacagatgctgtaagctcgattgaaggttataatgctgtgcttttgcacat 301  
QY 320 acggtcgacccgttcgggcaagtcctatctaatgataatgaggttataatgagagcatggcg 379  
Db 302 atgggcaactggatcggaaagtcacacactatgaggaattctggagattctggct 361  
QY 380 tgaatccgcgattgccaggacatgttccggcgatttaataatgaactcagaagagacaaga 439  
Db 362 taatcacctcgatctgtaaggactcttcagtcagtgataaatgaaccaccacagatgggatg 421  
QY 440 acctcaactgcacccgtcgaaatttcgtacttggaaatttacaacaattgtatcttcgct 499  
Db 422 aagcttttttcgaactgaagtcagctacttagaatttataacgaacgctgtgagagatc 481  
QY 500 t-----gctgaatccgtcgacaaaaggggaatctcaaggctccgagacacccgtcgaccg 553  
Db 482 tacttcggcggaagtcactcaaaccttcaatttgaggtcaggacacccagatgggatg 541  
QY 554 gccctacgtggagacttggcgaagctgctgctgcatcattcccaagaagaatcgaaatc 613  
Db 542 gccctattgttgagatttaccacaaacatttagtacagaattatggtgacgtagaagaac 601  
QY 614 tcattgagtgagggaacaaagccagaacggttgcgcgcaacaaatgaacgagacatcca 673  
Db 602 tatggatgcgggcaatatcaaccggaaccacgcagcagctgggtgaacgagctcagta 661  
QY 674 gtcatccacgcgcgttttcaacttgcattgacgcgcaaaagtggtcagatgagagaccca 733  
Db 662 gcaggtcatgcatcttccacatcaagttcaactca-----ggctaaaatttgattctg 715  
QY 734 aaatggcacacagagaaggttgcgaagatcagctgaaggagggtgcagagatcaaacgcgtcac 793  
Db 716 aaatggcatgtgaacccgtcagtaagatccacttgggtgattcttgcgggaagtgcgctg 775  
QY 794 aacgcgtccacggagctacttggagcgcgactgaaggagggtgcagagatcaaacgcgtcac 853  
Db 776 cagatgcacccggagccaccgggttaggtctaaagggaaggggaaatattaaacagtcgcc 835  
QY 854 tttagacctagctgctgtgattgacgctagcggatagcggtatgtcgttcgggaa----- 904  
Db 836 ttgtgactctggggaacgctcatttctccttagctgatttatctcaggtgctgcaataa 895  
QY 905 -----aacagaagaagaatcagttagcttaccgagattcgggtactcagctggcttc 958  
Db 896 ctcttgcaagaagaagaacagtttctgcttaccagggattctgtgtgacttgggtgtg 955  
QY 959 tgaaggactccttgggaggcaactcgatgaccgccatgattgcgcgcattttcgcctgctg 1018  
Db 956 taaaagatagcttgggaggaaactctaaactataatcatcattgattgcacatttcaactgctg 1015  
QY 1019 atattactttgaagagactctcagtaaccttgatagtcgggactctgcgaagcgaatca 1078  
Db 1016 atgccaattatgagaacccctaaagtactctctgctatgcaaatagagccaaaacatca 1075  
QY 1079 agaaccacgcagtggttcaatgaagacccgcagccgcggtatgatccgcgagttgaagagg 1138

Db 1076 tcaaaagcctaccattaatgaggatggcacaacgtcaaaacttatccgtgagctgcgagctg 1135  
QY 1139 aactcgcgcagctga 1153  
Db 1136 aaatagccagactga 1150  
RESULT 4  
US-09-722-139-3  
; Sequence 3, Application US/09722139  
; Patent No. 6355471  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1055  
; CURRENT APPLICATION NUMBER: US/09/722,139  
; CURRENT FILING DATE: 2000-11-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1077  
; TYPE: DNA  
; ORGANISM: Human  
US-09-722-139-3

Query Match 11.2%; Score 264; DB 4; Length 1077;  
Best Local Similarity 55.9%; Pred. No. 7.6e-69;  
Matches 604; Conservative 0; Mismatches 440; Indels 36; Gaps 4;

QY 20 tcaagtggttggtgcgggtacgcccgttcaacgcccgcagaaatcgacggtgcgcaaaat 79  
Db 11 tcaagtggttcggtgaggttcgcccattgacgaggaagagacttggaggccaagt 70  
QY 80 gtattgtcgcatggaaggaaatcaaacctctccaccctccctccgggtgcggaagaga 139  
Db 71 tcattattcagatggagaagaacaaacgacaaatcaaaacttaaaatataccagaag--- 127  
QY 140 aggcgcgtaaagtggcaaaactattatggatggcccggaaggcatttgcgttcgcatcgg 199  
Db 128 -----gaggcacttgggactcaggaagagacggaccaagaccttcacatgactttt 181  
QY 200 cgtattgtcctttgacaagaatgctcccaactatgacgacaggaagacactattccaag 259  
Db 182 cttttattctgctgatacaaaaagccagatcactgatttcacaagaagaatggtttcaaaa 241  
QY 260 atctcggagtcgcccttctgataatgcatcgaagggttataaacaattgtatcttcgct 319  
Db 242 ccttcggcacagatgctgtaagctcatttgaaggttataatgctgtgcttttgcacat 301  
QY 320 acggtcgacccgttcgggcaagtcctatctaatgataatgaggttataatgagagcatggcg 379  
Db 302 atgggcaactggatcggaaagtcacacactatgatactgatacttgcgggaattctgcggat 361  
QY 380 tgaatccgcgattgccaggacatgttccggcgatttaataatgaactcagaagagacaaga 439  
Db 362 taatcacctcgatctgtaaggactcttcagtcagtgataaatgaaccaccacagatgggatg 421  
QY 440 acctcaactgcacccgttcggaagtttctgacttggaaatttacaatgaacgagtgcgagact 499  
Db 422 aagcttttttcgaactgaagtcagctacttagaatttataacgaacgctgtgagagatc 481  
QY 500 t-----gctgaatccgtcgacaaaaggggaatctcaaggctccgagacacccgtcgaccg 553  
Db 482 tacttcggcggaagtcactcaaaccttcaatttgaggtcaggacacccagatgggatg 541  
QY 554 gccctacgtggagacttggcgaagctgctgctgcatcattcccaagaagaatcgaaatc 613  
Db 542 gccctattgttgagatttaccacaaacatttagtacagaattatggtgacgtagaagaac 601  
QY 614 tcattgagtgagggaacaaagccagaacggttgcgcgcaacaaatgaacgagacatcca 673



Db 602 ttaaggatgcgaggaataacacgagaccacgagcagcagactgggagacgacgtcagta 661  
Qy 674 gtcgatccacgacgcgtctcttaccatttgaccttgacgcaaaagtggcatgatgaagagacca 733  
Db 662 gcaggtctcatgccaactctccaccatcaagttcactca-----ggctaaattgattctg 715  
Qy 734 aaatggacacagagaaggttgcgaagatcagctcgtgtagatttggcgggttctgagcgag 793  
Db 716 aaatgccatgtgaacccgtcagtaagatccacttggttgatcttgccggaagtgcgctg 775  
Qy 794 caagtcacacgagcactactgagcgcgacacgaagaggggtgcagagatcaacgcgtcac 853  
Db 776 cagatgcaccgagccaccggggttagcctaaagggaagggaataattacaagtccc 835  
Qy 854 tttagaccctaggtgctgtagtgagcgcgttagcggatgctcgtcg------ 901  
Db 836 ttgtgactctgggaacgtcattctgcttagctgattatctcagatgctgcaata 895  
Qy 902 ---gaaacacagaagaatcagttagtagtacctaccgagattcggtagctgagcttc 958  
Db 896 ctcttgcaagaagaagcaagtttctgcttaccagggattctgtgtgacttggtgt 955  
Qy 959 tgaaggactcttgaggagcaactcgatgacgcgcacatgattgcgcacatttcgcctgctg 1018  
Db 956 taaagatagccttggagggaactctaaactatcatatgattgccaaccatttcacactgctg 1015  
Qy 1019 atattaacttgaagagactctcagtagccttctcgatgactcgtcgaagcgaaatca 1078  
Db 1016 atgtcaattatggagaacccttaagtactcttcgtatgcaaatagagccaaacatca 1075

RESULT 5

US-08-006-676B-2  
; Sequence 2, Application US/08006676B  
; Patent No. 5411865  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven  
; TITLE OF INVENTION: Diagnosis of Leishmaniasis  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jeffrey B. Oster  
; STREET: 8339 SE 57th Street  
; CITY: Mercer Island  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98040-4906  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORD for Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/006,676B  
; FILING DATE: 15-JAN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oster, Jeffrey B.  
; REGISTRATION NUMBER: 32,585  
; REFERENCE/DOCKET NUMBER: REED-4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 232 7845  
; TELEFAX: (206) 236 0205  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3319 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna to mRNA  
; ORIGINAL SOURCE:  
; ORGANISM: Leishmania chagasi  
US-08-006-676B-2

Query Match 7.6%; Score 177.8; DB 1; Length 3319;  
Best Local Similarity 52.6%; Pred. No. 7.3e-43;  
Matches 525; Conservative 0; Mismatches 417; Indels 57; Gaps 4;

Qy 214 gacaagaatgtctccaaactatgcgagacaggaagacctattccaaagatctcgaagtcocg 273  
Db 710 GACGGTCCGGCGCACCCCGCAGCAGGAGCAGCGTTCCTCCGACGATCCGGTATCCCG 769  
Qy 274 ctctggataatgcattcaagggttatacaattgtatctcttcgctcaggtcagaccgt 333  
Db 770 CTGTTGTCAGACGCGTTCACGGGTTCAACTCGTCTTGTGTTGCTACGGCGACAGAG 829  
Qy 334 tcgggaagtctctattcaatgatgggcta-----tggcaaggagcat 375  
Db 830 AGCGGGAAGACGTACACGATGATGGCGCGGAGCTGAGCGCGCTTAGTGGTGAGGGCAAC 889  
Qy 376 ggcgtgatcccgcgatttgcaggacatgttccggcgatttaataatgaactgcagaagac 435  
Db 890 GCGGTGACCGCGCGATCTGCTTGAGATCTTTCGCGGAAAGCGGAGCGTGGAGCGCAG 949  
Qy 436 aagaacctcacttgcacgctgaaagtcttctacttggaaattacaatgaacgagtgoga 495  
Db 950 GGGCACTCGCGGTGATCGTGAGCTGGGTACGTGGAGGTGTACACGAGCGCGTGTGCG 1009  
Qy 496 gacttgcgaatccgtcgacaaagg-----gaatctcaag 531  
Db 1010 GACCTGCTTGGGAAGCGGAAGAGGTGTGAAGGGCGCGGAGAGGTGTACGTGGAC 1069  
Qy 532 gtccgagaacaccgtgcaccggcccttacgttgaggagacttggcgaactggtcgtgoga 591  
Db 1070 GTGCGCGAGCACCCGCGCGGTGTTCCTGAGGGGACGCGCTGGTGGAGGTGGG 1129  
Qy 592 tcatccaaagaatcgaaatctcatgatgaggggcaacaaagccagaacggttccgcc 651  
Db 1130 AGCTGGAGATGTTGTGGGCTGATCGAGATCGGCAACGGCGTGGCGGCACACCGCTTG 1189  
Qy 652 acaaatgaacgagacatccagtcgacccgcgtctctcactttgacctt----- 704  
Db 1190 ACGAAGATGACACCGGAGCAGCGGAGCCAGCCAGCATCATCATCTGCTGCTGCGCGAG 1249  
Qy 705 ----gacgcaaaagtggcatgatgaagagacccaaatggacacagagaaggttgcgaag 759  
Db 1250 GAGCGGACGATGACAGCAGCAAGAGCGGGGAGACGATCCGTACTGCGCGCAAGAGACGCCGC 1309  
Qy 760 atcagctggttagatttggcgggttctgagcgacgaacgctccaccggaactactgagcg 819  
Db 1310 ATGAACCTTGTGGACCTTTCGGGGGTCTGAGCGCGTGGCGCAGTCCGAGTGGAGGGCGAG 1369  
Qy 820 cgaactgaaggaggtgcagagatacaaccgctcactttcgacctaggtgctgtgattgca 879  
Db 1370 CAGTTCAAGGAGCGACGCACATCAACCTGTCGCTGACGACGCTCGGGCGCTGTATCGAC 1429  
Qy 880 gcgctagcggatattgtctc---gggaaaaacagaagaagaatacagttagtaacctaccga 936  
Db 1430 GTGCTCGCGACATGGCGACGAAGGGTGGAAAGGGCGCAGTACAGCGCTTCGCGCGTTCGCC 1489  
Qy 937 gattcgtactgaactggtcttctgaagactccttggagggcaactcgatgacccgcatg 996  
Db 1490 GACTCGAAGCTGACGTTTCATCTCTGAAGGACTCGCTTGGCGGGAAGTCTCGAAGAGCTTCATG 1549  
Qy 997 attgcgcatttcgctcgtgtagatattaaacttgaagagactctcagtagtaccctcgat 1056  
Db 1550 ATCGCGACTGTGAGCCGAGCGCGTGAATACGAGGAGACGCTGAGCACGCTGCGGTAC 1609  
Qy 1057 gcggactctgcgaagcgaatcaagaacacgcagtggtcaatgaagaccgcgaacgcgag 1116  
Db 1610 GCGTCGCGCGCGGACATTTGTAATTGTTCGCGAGGTGAACGAGGACCCCGCGCACGG 1669  
Qy 1117 atgatccgcgagttgaaggagaaactcgcgcagctgagg 1155  
Db 1670 CGGATCCGCGAGCTGGAGGAGCAGATGGAGGACATGCGCG 1708





Qy 214 gacagaatgtctccaaactatcgagacagaaagacacctattcccaagatctcggagtgccg 273  
Db 710 GACGGTGCggCGGACCCCGCGAGCGAGCGAGCGATGTTCCGACCATCGGTACCCG 769  
Qy 274 cttctgataatgatattcaagggttataaacaattgtatcttcgactcaggtcgacacgggt 333  
Db 770 CTGTGTCAGCAGCGGTTTCGACGGGTTCAACTCGTGTGTTTGGCTACGGGCGACAGGG 829  
Qy 334 tcgggaagtccttcaatgatgggcta-----tgcaagagagcat 375  
Db 830 AGCGGAGAGCTACAGATGATGGCGCGGACGTGAGCGGCTTACTGTGAGGGCAAC 889  
Qy 376 ggcgtgatcccgcgagattgcccagacatgttcggcggtatttaatagaactgcagaagac 435  
Db 890 GCGGTGACGCGCGGATCTGCTGGAGATCTTTCGCGGAAGCGAGCGTGGAGGCGCAG 949  
Qy 436 aagaacctcactgcacgctcgaagtgttcgacttgtaaaatttaatagaacgagtgca 495  
Db 950 GGCACATCGCGGTGGATCGTGGAGCTGGGGTACGTGGAGGTGTACACAGCGCGTTCG 1009  
Qy 496 gacttgctgaatccgtcgacaaagg-----gaatctcaag 531  
Db 1010 GACCTGCTGGGAAGCGGAAGGGTGTGAGGGCGCGCGAGGAGGTGTACGTGGAC 1069  
Qy 532 gtccgagaacaccccgtaacggccctacgtggaggaacttggcgaagctggtcgca 591  
Db 1070 GTGCGGAGCAGCAGCGCGCGGTGCTGCTGAGGGGAGCGGCTGGTGGAGGTTGGG 1129  
Qy 592 tcattccaagaatacgaataatcctatgatgagggcgaacaaagccagaacggttgcgc 651  
Db 1130 AGCTTGACGATGTTGCGGCTGATCGAGATCGGAACGGCGTGGCGACACACGCTTCG 1189  
Qy 652 acaaacatgaacagagacatccagtcgatcccaagcgctctcactttgacct----- 704  
Db 1190 ACGAAGATGAACGACCGAGCAGCGCGAGCCACGCGATCATCTGCTGCTGCGCGAG 1249  
Qy 705 -----gacgcaaaagtggcatgataagagagacaaatggacacagagaggttgcgaag 759  
Db 1250 GAGCGGACGATGACGAGCAAGCGGGGAGACGATCGTACTGCCGCAAGAGCAGCGCG 1309  
Qy 760 atcagctgtgtagatttggcggttctgagcgagcaacgtccaccggagctactgaagc 819  
Db 1310 ATGAACCTTGTGGACCTTGGGGGTCTGAGCGGTGCGCGAGTCGAGGTGGAGGGCGAG 1369  
Qy 820 cgactgaagaggggtgcagagatacaacccgtcaacttcgacctaggtgctgtgattgca 879  
Db 1370 CAGTTCAAGGAGGCGCGCACATCAACTGCTGCTGACGACGCTCGGGCGCGTATCGAC 1429  
Qy 880 gcgctagcgatatgtcgtc-----gggaaaaaagaagaagaatcagttagttacctaccga 936  
Db 1430 GTGCTCGGACATGCGACGAAGGGTGGAGCGCGAGTACAGCGTTGCGCGCTTCCCG 1489  
Qy 937 gattcgctactgagtggtcttgaagactccttggaggaactcgtgacgcgcattg 996  
Db 1490 GACTCGAAGCTGAGTTTCACTCTGAAGACTCGCTTGGCGGGAACTCGAAGAGTTCATG 1549  
Qy 997 attgcgcacttgcctgctgatatatttaactttgaagagactcctagttacctcgatat 1056  
Db 1550 ATCGCGACTGTGAGCCCGACGCGGCTGAACACTAGAGAGAGCGTGAACGCTGCGGTAC 1609  
Qy 1057 gcggactctgcaaggaatcaagaaccagcgagttgtaataagaacccgaacgcgcgg 1116  
Db 1610 GCGTCCGCGCGCGGACATGTGAAATGTTGCGCAGGTGAACGAGACCGCGCGCACGG 1669  
Qy 1117 atgatccgcaggttgaagagggaactcgcgcagctgagg 1155  
Db 1670 CGGATCCGCGAGCTGGAGGAGAGATGAGGACATCGG 1708

RESULT 9  
US-09-718-841-1  
; Sequence 1, Application US/09718841

; Patent No. 6333184  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: NO. 6333184el motor proteins and methods for  
; FILE REFERENCE: 1051  
; CURRENT APPLICATION NUMBER: US/09/718,841  
; CURRENT FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1689  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (48)...(111)  
; OTHER INFORMATION: n = a, t, c, or g  
; NAME/KEY: misc\_feature  
; LOCATION: (1171)...(1236)  
; OTHER INFORMATION: n = a, t, c, or g  
US-09-718-841-1

Query Match 7.5%; Score 177.4; DB 4; Length 1689;  
Best Local Similarity 52.3%; Pred. No. 6.6e-43;  
Matches 534; Conservative 0; Mismatches 451; Indels 36; Gaps 5;

Qy 178 aaggcatttgcgttcgactcgtgattgtctcttgacaaagaatgctcccaactatgcg 237  
Db 157 atggcatttgcgttcgactcgtgattgtctcttgacaaagaatgctcccaactatgcg 216  
Qy 238 agacagaagacatttccaaagatcgcgagtcgcgcttcttgataatgattcaaggg 297  
Db 217 tctcaagatgggttttccaggatttagggatgaagtagtctctggattgccaaggc 276  
Qy 298 tataacaattgtatcttcgctacggtcgacacggttcggcgcaagtctcttcaatgatg 357  
Db 277 tataacatatgctttttcttattgacacagagctctggaagacataacatctgtg 336  
Qy 358 ggctatggcaagagagatggcggtgaccccggtattgcccagagacatgttccgcggtatt 417  
Db 337 ggagccaccagcctctgtgtgggttgacacacgatatgtgagggctctcttcgcaggag 396  
Qy 418 atgaaactgcagaaggacaaacacccctcacttcgacgctgaaagtcttcgacttggaaatt 477  
Db 397 aaaga---ctgtgcctcactgccttctcctctgtagataaaaagaagtttctcagaatc 453  
Qy 478 tacaatgaacagagtcgagacttgcgtgaatccgctcgacaaaggggaa-----tctc 528  
Db 454 tataatgaacgggtggcggtatctgttgaagcaatctggtaaaaaagtcctataccctg 513  
Qy 529 aaggtccgagaacacccgctgcacgcccctacgtgagagacttggcgaagtggctgtg 588  
Db 514 cgggtcagggagcagcagatggggccctatgtcaaaagtttctcacaatgtagtt 573  
Qy 589 cgatcttccaagaatacgaataatctcagtgatggaggcaacaaagccagaacggttgc 648  
Db 574 accaattataagcaagtaatacctaactcttggaggaggggaattgcaaacagaaaccaagca 633  
Qy 649 gccacaaatgaacgagacatccagtcgaccccgcttcttacttgaacttgaactgacg 708  
Db 634 gccacccaatttcatgaggcagcagcagatccacgcccattttcagatcccaacag 693  
Qy 709 caaaagtgcgatgaagagacaaatggacacagagaaggttgcgagagatcagttctg 768  
Db 694 ca-----ggcaatcctggagaaacacccctctctgaaatggctagcagatcaacctt 747  
Qy 769 gtgatttggcggttctctgagcgagcaacgctccacgggagctactggagcgagactgaag 828  
Db 748 gtggacctagcaggcagcaagagcagatccca---gttactgtgaaggacgcagctgct 804

Qy	829	gaggggtgcagagatcaaccgcctcaactttctcgacccttaggtgcgtgtgattgcagcgtagcg	888
Db	805	gaagagagccaatatcaacagtcctcttgactctaggaaattgtcatctccaccttagcc	864
Qy	889	gatatgcccc-----gtcgggaaacagaagaatactagttagtaaccttac	933
Db	865	cagaactccccaaagtttctgcagcagctgcagagcctcaacagctcatcttatatcccatac	924
Qy	934	cgagattcgggtactgacgtggcttctgaaggagactcctctgggaggcaactcgatgacgcgc	993
Db	925	cgagactctgtgtgacctggctgctggaaggacagccttgaggcaactctaaaccatc	984
Qy	994	atgattgcgcgaatttcgctgctgatatctaactttgaagagactctcgaatgacccttcga	1053
Db	985	atggttgccacgggtgctctctgcacacactagctacagtgcagcaatgagcacactgaga	1044
Qy	1054	tatgcggaactctcggaagcaatcaagaaccacagcagtggtcfaatgaagaccgcgaacgcg	1113
Db	1045	tatgcatactcagtgcccaaaaacattatcaacagccacagcaaatgaggatgcgaactta	1104
Qy	1114	cgatgatccgcgagttgaaggaggaactcgcgcagctgaggagcgaactccagacagt	1173
Db	1105	aaactgattagaaactcagagaagagattgaaagactgaaagccctgctgtgagcttt	1164
Qy	1174	g 1174	
Db	1165	g 1165	
RESULT 10			
US-09-718-841-3			
; Sequence 3, Application US/09718841			
; Patent No. 633184			
; GENERAL INFORMATION:			
; APPLICANT: Beraud, Christophe			
; APPLICANT: Freedman, Richard			
; TITLE OF INVENTION: NO. 633184el motor proteins and methods for			
; TITLE OF INVENTION: their use			
; FILE REFERENCE: 1051			
; CURRENT APPLICATION NUMBER: US/09/718,841			
; CURRENT FILING DATE: 2000-11-22			
; NUMBER OF SEQ ID NOS: 4			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 3			
; LENGTH: 1070			
; TYPE: DNA			
; ORGANISM: Human			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (48)...(111)			
; OTHER INFORMATION: n = a, t, c, or g			
US-09-718-841-3			
Query Match 6.7%; Score 158.2; DB 4; Length 1070;			
Best Local Similarity 52.7%; Pred. NO. 2.8e-37;			
Matches 483; Conservative 0; Mismatches 398; Indels 36; Gaps:			
Qy	178	aaggcatttcgctcgatcggtcgctattggtctcttgacagaagaatgctcccaactatgcg	237
Db	157	atggtcatttgccttctgattactgctactggtgcagtcacccagaggatccccagatgca	216
Qy	238	agacaggaagacctattccaagatctcgagtcgccgctctctggataatgcattcaagggt	297
Db	217	tctcaagatgggtttctccagatttagggatggaagcactgctcgtgagttgcgaaggcc	276
Qy	298	tatacaattgtattctcctacgctcagcagccggttcctggcaagctctattccaatgatg	357
Db	277	tatacatatgccttttcttatgatgacagacagcgtctggaagacatataccatgctg	336
Qy	358	ggctatggcaaggagcatggctgtgatacccgcggaatttgcaggagcaatgttccggcggtatt	417
Db	337	gggacccagcctctgttgggtttgacaccacagatgattgaggtctctctctcagaggaq	396



; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-713-815A-2

```
Query Match          4.5%; Score 107.8; DB 2; Length 3572;
Best Local Similarity 49.9%; Pred. No. 5.8e-22;
Matches 444; Conservative 0; Mismatches 397; Indels 48; Gaps 5;

QY 233 atcgagacaggaagaccattccaaagatcgcgagtcgcgcgttcctggataatgcatcca 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 505 ATGCATCCCAAGAAAGGCTACAAATGAGCGGCCCAAGTCCATGTTACGGATGTCCTGG 564

QY 293 aggggtataacaattgtattcttcgcctacggtcagaccggttcgggcaagctattcaa 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 CCGGCTACAATGAACGATATTCGCATATGCTCAGACGTCCTCCGGAACACATACGA 624

QY 353 tgatggctatggcaagg-----agcatggcgtatcccgcgagattggcaggaca 403
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 625 TGGAGGGCGTGATCGGGGACTCCGTAACACAGGGTATCATACCACGTATCGTCAACGACA 684

QY 404 tgttcggcggtattaatgaactcagaagacaaagaccctcacttgccacgctcgaagttt 463
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 685 TTTTCATCATCTACG-----CGATGGAGGTGAACCTGGAGTTTCACATCAAGGTCT 738

QY 464 cgtacttgaaattacaatgaacgagtcgagacttgcgtgaatccgctgcacaaaggga 523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 739 CCTACTACGAGATCTACATCGCAAGATTTCGAGATCTGTTGGACGTCT---CCAAGGTGA 795

QY 524 atctcaaggtccgagaaacccgtgcacccgccccctacgttgaggacttggcgaaagctgg 593
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 ACCTTAGTGTGCACGAGGATGAAGAACCGGTCGCCGTACGTCACGTCACGAGCGGTAC 855

QY 584 tcgtgcgcatctccaaagaatcgaaatctcatgtagtggggcgaacaaagccagaacgg 643
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 TCGTCTCGTCGCGGAGGATGTTTTCGAGGTGATCGAGGGGCAATCAATCGTCACACA 915

QY 644 ttgcgcgcaacaacatgaacgagacatccagtcgatcccaacgcgcttcactttgacct 703
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 TCCTCTGTGACAAACATGAACGAGCATTTCTTCGCGATCCCACTCAGTATTTCTTATCAATG 975

QY 704 tgacgcaaaagtggcatgatgaagacacaaatggacacagagaaggttgcgaagatca 763
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 976 TGAAGCAGGAGAACCTGGAGAACCAAGAAATACTATCC-----GGCAAACTCT 1023

QY 764 gtctgtagatttggcgggttctgagcgagcaacgtccaccggagctactggagcgcgac 823
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1024 ACCTGTGGATTTGGCCGTTTCGAGNAGGTTTCCNAGACTTGGCGGGGGAACCGTTC 1083

QY 824 tgaaggaggggtgcagagatacaaccgtcacttctcgaccctagtgctgtgattgcagcgc 883
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1084 TTGATGAAGCCAAAGAACATCAACAAGTCGGTGTTCGGCCTTTGGGCAACGTAATTTCTGCC 1143

QY 884 tagcgatatgtcgtcgggaaacagagaagaatcagttagttacattaccgagattcgg 943
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1144 TGGCGGACGGA-----AACAAACGCAATCTCCCTACCGTGATTTCCA 1185

QY 944 tactgcgtggcttctgaaggactccttgggagggcaactcgtgaccgcatgatgtgcgg 1003
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1186 AGTAACGGCGATCTTCGAGGAGTCGTGGGAGGCAACGCACGCAACCATCGTCATCT 1245

QY 1004 ccaatttcgctgctgatatataatttgaagagactctcagtaacccttcgatatgcggact 1063
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1246 GCTGCTCTCCAGCCAGTTTCAACGAAATCTGAAACGAAGTCAACGCTGAGACTTCGGTGGTA 1305

QY 1064 ctgcgaaggaatcaagaaccacgagtggtcaatgaagacccggaacgc 1112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1306 GAGCCAAGACAGTGAAGACGTGCTTCGCTTTAACGAGGAGCTTACTGC 1354
```

```
; Sequence 5, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: NO. 6355466el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Human
US-09-572-191-5
```

```
Query Match          4.5%; Score 106.4; DB 4; Length 1230;
Best Local Similarity 51.1%; Pred. No. 8.4e-22;
Matches 338; Conservative 0; Mismatches 306; Indels 18; Gaps 3;
```

```
QY 462 ttctgacttgaaattacaatgaacagagtcgagacttgcgaacttcgaatccgtcgacaaaggg 521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 tctctttatgaaatctacaacgagcagatatatgatctactgga---ctctgcacggc 548

QY 522 gaatctcaaggttcgagaaacacccgtcgacccgcccctagctggagacttggcgaaagct 581
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 549 tggactgtacttaaggaggacatacaagaaggaggagttcttgtgtgtgtgtgtgtgtgtgtgt 608

QY 582 ggctgctgcgtatctccaaagaatcgaataatcctatgcgtgagggagcacaagaagccagaac 641
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 609 ggtggtaacctcagctgctgaagcctatcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 668

QY 642 ggttgcgcgcacaaacatgaacgagacatccagtcgagtcgacccacgctcttcactttgac 701
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 669 tgtggcatcaactcaatgaacgagagatcgtctagctctcatcgtctcttttacaattac 728

QY 702 ctgacgcgaaggtggcatgatgaagagacccaataatggacacagagaggttgcgaagat 761
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 729 aatagatcaatggagaaaagtaagtga-----atgtgaatacaggaacctccctact 782

QY 762 cagtcgtgtagatttggcgggttctgagcgagcaacgtccaccggagctactggagcgcg 821
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 783 caacctgtgtgatttagcaggatctgaaaggcaaaagatacccatgcagaagggatgag 842

QY 822 actgaagagaggttgcagagatcaaccgctcactttcgaccctaggtcgtgtgattgcagc 881
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 843 attgaaggagcaggttaacataatcgatcattgagctgctgctggcgaagtattacagc 902

QY 882 gctagcggtatgtcgtcgggaaaacagagaagaatacagttagttacttaccgagattc 941
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 903 actgtcgaagtggttaataaggaaaacagagaca-----tgttgcacagagactc 953

QY 942 ggtactgacgtggtcttgaaggactcctctggggaggaactcgtatgcccgcatgattgc 1001
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 954 caaacttaaccttctactacgggattcccttggaggtaatgccaaaacagcacaattgc 1013

QY 1002 cgccatttcgctgctgatatataactttgaagagactctcagtaaccttcgatatgcgga 1061
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1014 aaatgttcaacctctggatccaggtgttttgggaaacctatcaacacttctgtcga 1073

QY 1062 ctctgcgaagcgaatcaagaaccacgagtcgaatgaagacccggaacgcgaggtgat 1121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1074 aagagcaagctgattaaaaaagcagtagtaataatgaagacaccccaaggaatgtgag 1133

QY 1122 cc 1123
   ||
Db 1134 cc 1135
```

Search completed: July 25, 2002, 05:38:29  
Job time: 6617 sec

```
RESULT 15
US-09-572-191-1
; Sequence 1, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4757
; TYPE: DNA
; ORGANISM: Human
US-09-572-191-1

Query Match      4.5%; Score 106.4; DB 4; Length 4757;
Best Local Similarity 51.1%; Pred. No. 1.8e-21;
Matches 338; Conservative 0; Mismatches 306; Indels 18; Gaps 3;

Qy 462 ttcgtaacttgaaattacaatgaacagagtcgagacttgcgaatccgtcgacaaagg 521
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 492 ttcctttattgaattcacaacagcagcatatgatctactgga---ctctgcacggc 548
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 522 gaatcacaaggtccgagacacccgtcgacccgcccctacgtgagagacttgggaagct 581
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 549 tggactgtacttaaggagagatatcaagaaggaggtctcttgttggcggtggagca 608
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 582 ggtcgtgcgatactccaagaatactcaaaatctcatgagtgaggggcaacaagccagac 641
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 609 ggtgtaacctcagctgtcgaagctatcaggtgttctgtgagagtgaggaatagacg 668
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 642 ggttgcgcacaacaatgaacagagacatccagtcgatacccaacgcgtcttcaactttgac 701
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 669 tgtggcatcaacatcaatgaacagagagatcgtctaggtctcatgctcttacaattac 728
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 702 cttagcgaagaagtggcatgatgaagaccaaataatgacacagagaaggttgcaagat 761
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 729 aatagagtcaatggagaaaaagtaatgag-----attgtgaatatacggacctccctact 782
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 762 cagtcgtgtagatttggcgggttctgtgagcgagcaacgtccaccggagctactgagcgcg 821
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 783 caactggtggtattagcaggatctgaaaggcaaaagatacccatgcagaaggatgag 842
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 822 actgaaggaggtgcagagatcaaccgtcaacttgcacccctaggtcgtgtgattgcagc 881
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 843 attgaaggaaagcaggttaacataaatcgtcattgagctgctgggccaagtattacagc 902
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 882 gctagcggtatgtcgtcgggaaacacagaagaataatcagttagcttaccgagattc 941
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 903 actgtcgacgtgggtaattggaaacacagagaca-----tgtttgctacagagactc 953
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 942 ggtactgacgtggcttctgaaggactccttgggagggcaactcgatgaccgccatgttc 1001
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 954 caaacttaccttcttactacgggtaccccttggaggtaatgccaaaacagccataattgc 1013
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1002 cgccatttcgctctgataatttaacttgaagagactcagtagctacccttcgatatcgga 1061
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1014 aatgttcatcctggatcccggtgttttggggaacccctatcaacacttaactttgctca 1073
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1062 ctctcgaagcgaatcaagaaccacgcagtggtcaatgaagaccccgacgcgcggatgat 1121
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1074 aagagccaagtgtataaacaaggcagtagtaataatgaagacaccccaagaaatgtgag 1133
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1122 cc 1123
Db   ||
Qy 1134 cc 1135
Db   ||
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:46:42 ; Search time 2969.55 Seconds  
(without alignments)  
10690.117 Million cell updates/sec

Title: US-09-235-416-2  
Perfect score: 2352  
Sequence: 1 atgtcggcggtggaatat.....aaacagcgaagcgaattc 2352

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	13.6	1683	11 BC005977	BC005977 Homo sapi
2	277.4	11.8	1074	10 B1906040	B1906040 603062382
3	277.2	11.8	670	9 A1816019	A1816019 au44c12.x
4	270.2	11.5	1263	11 AK020905	AK020905 Mus muscu
5	264.4	11.2	694	10 B1281667	B1281667 UI-R-CT08
6	260	11.1	761	10 B1414465	B1414465 602986047
7	256.8	10.9	1014	10 B470467	B470467 AGENCOURT
8	255.6	10.9	853	9 A1363308	A1363308 qv57c04.x
9	255.6	10.9	853	9 A169473	A169473 AU169473
10	255.2	10.9	774	10 BG473130	BG473130 602515112
11	236.8	10.1	670	9 A1879422	A1879422 au50f08.x
12	233.4	9.9	592	10 BG297948	BG297948 602396138
13	233.2	9.9	564	9 A1570691	A1570691 tm79f05.x
14	226.4	9.6	586	10 BE252943	BE252943 601117534
15	225.8	9.6	670	10 BF468073	BF468073 UI-M-CD0-
16	224.6	9.5	481	9 A1948449	A1948449 wp90h08.x
17	223.4	9.5	690	10 B1685690	B1685690 603309531

C 18	222.4	9.5	507	9	A1339237	A1339237 qt06g12.x
C 19	220.6	9.4	634	9	BE164569	BE164569 RC3-HT047
C 20	220.6	9.4	932	9	A1323149	A1323149 mk16h12.y
C 21	220	9.4	1030	10	BM451369	BM451369 AGENCOURT
C 22	216	9.2	588	9	AV594757	AV594757 AV594757
C 23	215.8	9.2	479	9	A1360405	A1360405 qy85h09.x
C 24	213.2	9.1	561	10	BF150140	BF150140 uy82c07.y
C 25	210.4	8.9	458	10	BE950986	BE950986 UI-M-BH4-
C 26	210	8.9	570	10	BF921560	BF921560 MRL-NT017
C 27	208.4	8.9	456	9	AW125600	AW125600 UI-M-BH2
C 28	207.6	8.8	588	9	AW898094	AW898094 RC3-NN006
C 29	206.6	8.8	646	9	AA097269	AA097269 mk16h12.f
C 30	204	8.7	972	10	BM475082	BM475082 AGENCOURT
C 31	202.8	8.6	423	9	A1323139	A1323139 mk10e06.y
C 32	200	8.5	532	9	AU060260	AU060260 AU060260
C 33	198.4	8.4	485	10	BE861354	BE861354 UI-M-AM0-
C 34	198	8.4	660	10	BE258249	BE258249 601115912
C 35	197.2	8.4	457	10	BF548635	BF548635 UI-R-A0-a
C 36	195	8.3	494	9	A1933167	A1933167 wp64d10.x
C 37	193	8.2	799	10	BG393452	BG393452 602411816
C 38	192	8.2	869	10	BE380022	BE380022 601159336
C 39	190.2	8.1	443	9	AA097858	AA097858 mk10e06.f
C 40	189.8	8.1	489	10	BF281801	BF281801 EST446392
C 41	189	8.0	560	10	BG202093	BG202093 602822811
C 42	186.2	7.9	424	9	AA456792	AA456792 zw27h02.f
C 43	185.8	7.9	585	10	BE887181	BE887181 601508384
C 44	181.2	7.7	545	9	AW360994	AW360994 PM1-CT024
C 45	177.8	7.6	417	9	BB733392	BB733392 BB733392

ALIGNMENTS

RESULT	1
BC005977	BC005977
LOCUS	Homo sapiens, Similar to kinesin family member 13B, clone
DEFINITION	IMAGE:4103715, mRNA.
ACCESSION	BC005977.1 GI:14710838
VERSION	BC005977
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 1683)
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	Strausberg, R.
TITLE	Direct Submission
COMMENT	Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabps-remail.nih.gov">cgabps-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: CLONTECH Laboratories, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a> Contact: (Dickson, Mark) <a href="mailto:mcdepaxil.stanford.edu">mcdepaxil.stanford.edu</a> Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 21 Row: d Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8896163 This clone has the following problem: frame shifted.
Location/Qualifiers	1..1683
source	



Matches 524; Conservative 0; Mismatches 326; Indels 11; Gaps 4;		ORGANISM	Homo sapiens
QY 311	tcttcgctacggtcagaccggttcgggcaagctcctattcaatgatggtgctatgccaagg 370	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Db 885	TTTTGGCCTAAGCCAGACTGTGTGGGAAATCTTAATACATGTGGTAAACAAGAAG 826	AUTHORS	1 (bases 1 to 670) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
QY 371	-----agcatggcgtatcccgcgagattgcccagacatgttccggcgatataatgaac 424	TITLE	WashU-NCI human EST Project
Db 825	AAACCCAGCGCTGGGATCCATCCACATTTATGTGAAGAACCTTTTAGAGAAAATCAATG 766	JOURNAL	Unpublished (1997)
QY 425	tcgagaagacagaacacacctcactgcaccgtcgaagttcgtacttggaatttacaatg 484	COMMENT	Other ESTs: au44cl2.y1 Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand Seq primer: -40UP from Gibco High quality sequence stop: 459.
Db 765	ACCACTGTAATCAAGAAATGCTTACTCTGTAGAGTGAGCTACATGG-AAATTACTGTG 707	FEATURES	Location/Qualifiers 1..670 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2517622" /clone_lib="Schneider fetal brain 00004" /sex="male" /tissue_type="frontal lobe" /dev_stage="5 months post-conception" /lab_host="DH10B"
QY 485	aacgagtgagactgtctgaatccgctcgacaaaggggaatctcaaggtccgagaaaccc 544	source	
Db 706	AAAGAGTACGAGATTTGCTGAATCCAAAAC-AAGGGTAATTTGCGTGTGGAACACC 648		
QY 545	cgtcaccgcccctacgtggaagacttgccgaagctggtcgtcgatcattccaaagaaa 604		
Db 647	CACTTCTTTGGACCCATATGTGGAGGATCTGTCCAAGTTGGCAGTTACTTCTTACACAGACA 588		
QY 605	tcgaaaaatctatgatgagggcaacaagccagacggttccgcgcacaaacatgaacg 664		
Db 587	TTGCTGACCTCATGATGCTGGGACAAAGCCAGGACAGTGCAGCTACAAACATGAATG 528		
QY 665	agacatccagtcgacccacgcgtctctcaattgacettgcgcgcaaaagtggcatgatg 724		
Db 527	AAACAAGTAGCGTTCACCGCTGTGTACGATTTTACGATTTTTCACCCAGAAAGAACAGATA 468		
QY 725	aagacacaaaatgacacagagaaggttcgaagaatcagtcagttgtagattggcggtt 784		
Db 467	ATGAGACCAACCTTTCCACTGAGAAGGTGAGTAAATCAGTTGTTGGATCTAGCAGGAA 408		
QY 785	ctcagcaggaacagctccaccgagctactggagcgcgactgagagaggttgacagatca 844		
Db 407	GTGACGAGCTGATTCACACTGTGTCACAAAGGACTCGATTAAAGNAGGAGCAATATTA 348		
QY 845	accgtctactcttcgaccctagctgtgtgattgcagcgtcgtcgatattgctgtgggaa 904		
Db 347	ATAAGTCTCTCAACATTTGGGCAAGTCAATTTTCAGCCTTGGCCGA--GGTCAAGTAAAA 291		
QY 905	aacagaagaagaatcagttagctaccgagattcgtactgacgtggtgcttctgaagg 964		
Db 290	AGAAGAAGAAACAGATTTATTCCTACAGGGATTTCTGTACTTACTTGGCTCCCTTCGAG 231		
QY 965	actccttgggagcgaactcgatgcccgccatgattgcgcctatttgcctgctgatca 1024		
Db 230	AAATTTAGTGGCAATTCCTCGACTGCATGTTGCTGCTCTGAGCCCGCGGATATCA 171		
QY 1025	acttgaagagactctcagtcaccttcgatatcggaactctcggaagcgaatacaagaacc 1084		
Db 170	ACTACGATGAGACTTTGAGCACTCTGAGATATGCAGATCGTGCAAAACAAATTAATGCA 111		
QY 1085	acgcagtggttcaataaagaccgaacgcgcggaatcccgagattgaaggaggaactcg 1144		
Db 110	ATGCTGTTATCAATGAGGACCCCAATGCCAACTGCTCGTAATTAAGAGGAGGAGTGA 51		
QY 1145	cgcagctgagggaacactcc 1165		
Db 50	CACGCTGAAGGACCTTCTTC 30		
RESULT 3			
AI816019/c			
LOCUS	AI816019	670 bp	mRNA linear EST 09-JUL-1999
DEFINITION	au44cl2.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2517622 3' similar to SW:KFLA_HUMAN Q12756 KINESIN-LIKE PROTEIN KIF1A ; mRNA sequence.		
ACCESSION	AI816019		
VERSION	AI816019.1	GI:5431565	
KEYWORDS	EST.		
SOURCE	human.		

Query Match	11.8%;	Score 277.2;	DB 9;	Length 670;
Best Local Similarity	65.7%;	Pred. No. 2.2e-63;		
Matches 437;	Conservative 0;	Mismatches 219;	Indels 9;	Gaps 2;
QY 259	gatctcgagtcggttcgttgataatgcatcgaagggttatacaaatgtatcttcgcc 318			
Db 667	GACATCGGAAGNAGATGCTGCAGCATGCTTTGAGGGATACACAGTGTGCATCTTGGCC 608			
QY 319	tacggttcagaccggttcgggcaagctcttcaatgatgggctatggcaag-----gag 372			
Db 607	TATGGCAGACGGGTGCGGCAAGTCTTACACCATGATGGCAAGCAGGAGAGGACCAG 548			
QY 373	catggcgtgatcccgcggtatttccaggagcatgttccgcgctattaatgaactgcagaag 432			
Db 547	CAGGGCATCATCCACAGCTCTCGAGAGCCTCTTCTCTCGATCAACGACACACCAAC 488			
QY 433	gacaaagacactcacttgaccgtcgagttctgacttgaaatttacaatgaacgagtg 492			
Db 487	GAC---AACATGTCTTACTCCGTGGAGGTGAGTACATGAGAGATTACTGTGAGCGGTC 431			
QY 493	cgcagacttgtgaatccgtcgacaaaggggaatctcaaggtccgagaaacacccgtcgacc 552			
Db 430	CGTGACCTCTTGAGACCCCAAGACACAGGGCAACCTTCGCGTGAGGGAGCACCACTGCTG 371			













```

ACCESSION   BG473130
VERSION     BG473130.1  GI:13405405
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 774)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L1Cm1422 row: h column: 04
            High quality sequence stop: 753.
FEATURES    Location/Qualifiers
             1..774
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:4646835"
             /clone_lib="NIH_MGC_16"
             /tissue_type="retinoblastoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
             EcoRI; cDNA made by oligo-dT priming. Directionally
             cloned into EcoRI/XhoI sites using the following 5'
             adaptor: GCGACGAG(G). Library constructed by Ling Hong
             in the laboratory of Gerald M. Rubin (University of
             California, Berkeley) using ZAP-cDNA synthesis kit
             (Stratagene) and Superscript II RT (Life Technologies).
             Note: this is a NIH_MGC Library."
BASE COUNT  203 a 237 c 213 g 121 t
ORIGIN

Query Match      10.9%; Score 255.2; DB 10; Length 774;
Best Local Similarity 65.3%; Pred. No. 2e-57;
Matches 440; Conservative 0; Mismatches 223; Indels 11; Gaps 4;

Qy 228 caactatgcgagacaggaagacctattccaagatctcggagtcgcggtctctgataatgc 287
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 63 CAAC*ACGCGTCGCAGAGCAGGTGTACCGGACATCGCGGAGGAGATGCTGCAGCATGC 122

Qy 288 attcaaggggtataacaattgtattcttcgctacaggtcagaccggttcgggcaagtcta 347
    || |||| ||| ||| ||||| ||||| ||| ||||| ||||| ||| ||||| |||||
Db 123 CTTTGAGGGATACAACGGTGTGCATCTTCGCTATGCGGACAGCGGTGCGGCAAGTCCTA 182

Qy 348 ttcaatgatgggtctatggaag-----gagcatggcgtgatcccgcggtattgcaggaa 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CACCATGATGGCAGCAGGAGAGAACGACGACGAGGCGATCATCCACACCTCTGCGAGGA 242

Qy 402 catgttcgcggcgattaatgaactgcagaagagacaagaacctcaattgcaccgtcgaagt 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 CCTCTTCTCTCGGATCAACGACGACGACCAACGAC---AACATGTCTACTTCGCTGGAGGT 299

Qy 462 ttctacttgaaatttacaatgaacagatgcgagactgtgactggaatccgtcgacaaagg 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 CAGCTTACATGAGATTTACTGTGAGCGGTGCGGTGACCTCTTGAACCCCAAGAACAGGG 359

Qy 522 gaattctcaaggtcccgagaaaccccgtagcccgccctacgtggaggagcttggcgaagct 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 CAACCTTCGCGTAGGAGCACCACCTGCTGGGGCCCTACGTGGAGGACCTCTCCAAGCT 419

Qy 582 ggtcgtgcgatactccagaataatcgaataatctcgtgatggggcaacaaagccagaac 641
    || |||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 420 GCGTCTCACCTCCTACATGACATCCAGGACCTCATGGACTCAGGGAACAAGGCCAGGAC 479

```

---

```

Qy 642 ggttcgcccacaaacatgaacgagacatccagtcagatcccccgcgcttctcactttgac 701
    || || || || || || || || || || || || || || || || || || || || ||
Db 480 CGTGG-CGCGCACCAACATGAATGAGACCAGCATGCTCCACGCGCTTCAACATCAT 538

Qy 702 cttagcgcaaaagtggcatgatgaagagaccacaaatggacacagagaaggttgcgaagt 761
    ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 CTTACCCAGAGCCCATGACGACGAGACCAATATCACCACGGAGAGGTGAGCAAAAT 598

Qy 762 cagttggg-tagatttggcgggttcttgagcgagcaacgtccaccggaggtacttgagcg 820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 CAGCCTGTTGGACCTGGCTGGGAGCGAGCGGCTGACTCCACGGGAGGACGACGC 658

Qy 821 gactaaagaggggtcagagatcaaacccgctcaatttcgaccctaggtctgtgattgag 880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 GCGTCAAGAGGGGGGCCAACATCAACAGTCGTCGACGAGCTGGGCAAGGTCATCTCCA 718

Qy 881 cgttagcggtatag 894
    || || || || || || || || || || || || || || || || || || || || ||
Db 719 GCGTGGCTGAAATG 732

RESULT 11
LOCUS     AI879422/c
DEFINITION au50f08_x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
            IMAGE:2518215 3', similar to SW:KFIA_HUMAN Q12756 KINESIN-LIKE
            PROTEIN KIF1A ;, mRNA sequence.
ACCESSION AI879422
VERSION   AI879422.1  GI:5553471
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 670)
            Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wyllie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
            Other ESTs: au50f08.y1
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -400P from Gibco
            High quality sequence stop: 423.
FEATURES    Location/Qualifiers
             1..670
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:2518215"
             /clone_lib="Schneider fetal brain 00004"
             /sex="male"
             /tissue_type="frontal lobe"
             /dev_stage="5 months post-conception"
             /lab_host="DH10B"
             /note="Organ: brain; Vector: pBluescript SK (Stratagene);
             Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
             prepared from human fetal brain tissue. 5' and 3'
             adaptors were used in cloning as follows: 5' adaptor
             sequence:
             5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATTCCTCCCTCCCCCCC-3'
             and 3' adaptor sequence:
             5'-GAGAGAGAGACTCGAGTGTGTTTTTTTTTTTTTTT-3'. The library was
             size-selected for >0.5 kb inserts and has an average
             insert size estimated at 1.2 kb. This library was

```

constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy)."

BASE COUNT 129 a 186 c 168 g 186 t 1 others

ORIGIN

Query Match 10.1%; Score 236.8; DB 9; Length 670;  
Best Local Similarity 63.4%; Pred. No. 1.6e-52;  
Matches 413; Conservative 0; Mismatches 228; Indels 10; Gaps 3;

QY 273 gctctgataatgcattcaagggtttatacaattgtatcttcctcagtgacagcg 332  
Db 656 GATGTCGACGAATCTTGTAGGGATAACATGTGTGCATCTCGTTAATGCAAGGCG 597

QY 333 ttcgggcaagtcctattcaatgaggtctgcaaggagcat-----ggcgtgaccc 386  
Db 596 TGCTGGTANGCTTATTATGATGGGAMAGGACGAGGATTAGTAGGGTATCATTC 537

QY 387 gcggatttcagagacatgttccggcggtattatgaactgcagaagacaaacctcac 446  
Db 536 AAGATTCTGCAGGAC-TCATTATCTGGATTACTGACAGCACTAATG--ATAACATGTC 481

QY 447 ttgcacgtcgaaatttcgtacttgaaatttacaatgaacgagtcgagacttgcgaa 506  
Db 480 TTATCTGTGGAGGTACGATACATGGAGATTTACTGTAGTGGCTGCGTCACTTCTTGA 421

QY 507 tccgtcgacaaaggggaatctcaagggtccgagaaacccgtcgaccggccctcagtcga 566  
Db 420 CCCCAAGAACAAAGGCAACCTTTGTGTAGGGACGCCCACTGNTGGGGCCCTACGTGGA 361

QY 567 ggaattggcgaagctgtctgcgtcattcccaagaatacgaaaaatctatgtagtgg 626  
Db 360 GGACTCTCCAAAGCTGTGTACCTCTTACATGACATCAGGACGCTCATGGACTCAGG 301

QY 627 caacaaagcagaacggttgcgcgcacaaacatgaacgacatccagtcgattcccaacg 686  
Db 300 GAACAAGGCCAGGACCGTGGGGCCACCAATGATGACACGAGTGCCTCCCAAGC 241

QY 687 cgtctcaatttgacctgacgcgaaggtgcatgtagaagagacacaaatggagacaga 746  
Db 240 CGTCTTCAACATCATCTTCAACCAAGAGGCCATGAGCAGAGACCAATATCACACGGA 181

QY 747 gaaggttcgaagatcagtcgttagattggcggttctgagcgagcaacgctccacgg 806  
Db 180 GAAGCTGAGCAAAATACGCTGTGTGGACCTGGCTGGGAGCGGCGGCTGACTCCACGGG 121

QY 807 agctactggagcgagctgaaggggtgcagagatcaacccgtcactttcgacctagg 866  
Db 120 AGCCAAGGCGACGGCTCAAGGAGGGGGCCAAACATCAACAAGTCTGACCACTGGG 61

QY 867 tcgtgtattgcagcgtagcgatatgtcgtcggggaaacagaagaagaa 917  
Db 60 CAAGGTCACTCCGCCCTGGCTGAAATGGACTCCGGACCCCAAGAACA 10

RESULT 12  
BG297948 592 bp mRNA linear EST 21-FEB-2001  
LOCUS 602396138F1 NIH\_MGC\_94 Mus musculus cdna clone IMAGE:4507905 5',  
DEFINITION mRNA sequence.  
ACCESSION BG297948  
VERSION BG297948.1 GI:13062110  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 592)  
TITLE NIH-MGC http://mgi.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHAM10385 row: k column: 10  
High quality sequence stop: 578.

## FEATURES

source Location/Qualifiers  
1..592  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4507905"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 175 a 172 c 147 g 98 t

## Query Match

9.9%; Score 233.4; DB 10; Length 592;  
Best Local Similarity 65.4%; Pred. No. 1.2e-51;  
Matches 358; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 371 agcatggcggtgacccggtgatttgcagacatgttccggcggtattataatgaactgcaga 430  
Db 34 AGCAAGGCATCATTCACACACTCTGCGAAGACCTCTCTCTCGAATCAAGATTCCACCA 93

QY 431 agacaagaacactcacttgcacccgtgcgaagtcttctacttggaaattacaatgaacgag 490  
Db 94 ACAGC---AACATGTCTCTCTGTGGAGGTGAGTACATGAGAGATTACTGTGAGCGAG 150

QY 491 tgcgagacttgcgtcgtcgcagaaaggggaatctcaaggtcccgagaaacccctcga 550  
Db 151 TCGTGACCTCTGACCCCAAGAACAAAGGCAACCTGCGTGTGAGGGAGCACCCCTCTGC 210

QY 551 cggccctctacgtgggagacttggcgaagctggtcgtgcgcatcttccaagaatacgaaa 610  
Db 211 TGGGGCTTATGTGGAGGACCTCTCCAAAGTGGCTGTAACCTCTTACATGACATCCAGG 270

QY 611 atctcatgagtgagggcaacaaagcagaaacgggttgcgcccacaaacatagaacgacat 670  
Db 271 ACCTCATGACTCAGGGAAACAAGGCCAGGACCTGTGGCGGCCCAACACATGATGAGACCA 330

QY 671 ccagtgcattccacgcgcttctcacttgcaccttgacgttgacgcaaaagtgcgatgaagaga 730  
Db 331 GCAGCGTCTTACGCGAGCTTCAACATCATCTTTACCCGAGAGGCCCATGATGACAGAGA 390

QY 731 ccaaaatggacacagagaaggttgcgaagatcagctgtgtagatttggcgggttctctgagc 790  
Db 391 CCAACATCACACAGAGAAAGTGAAGCAAAATCAGCTGTGGACCTGGCAGCGAGTGAAC 450

QY 791 gagcaactccaccgagctactgcagcgcgactgaaggagggtgcagagatacaaccgct 850  
Db 451 GAGCCGACTCCACAGAGGCCAAGGGCACGCGACTCAAGGGGGGGGCCAACATCAACAAGT 510

QY 851 cacttgcacctagtgctgtgattgtgcagcgcgtacggtatgtcgtcggggaaacaga 910  
Db 511 CTTTGACCACTGGGACAGGTCATCTCTGCTGTGGCTGAATGACTCTGACCCCAACA 570

QY 911 agaagaa 917  
Db 571 AGAACAA 577

## RESULT 13

A1570691/c  
 LOCUS  
 DEFINITION  
 tm9f05.x1 NCI\_CGAP\_Brn25 Homo sapiens CDNA clone IMAGE:2164353 3', similar to SW:KFA1\_HUMAN Q12756 KINESIN-LIKE PROTEIN KIAF ;, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 A1570691.1 GI:4534065  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 564)  
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGA), Tumor Gene Index  
 Unpublished (1998)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 770 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 293.  
 Location/Qualifiers  
 1..564  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2164353"  
 /clone\_lib="NCI\_CGAP\_Brn25"  
 /tissue\_type="anaplastic oligodendroglioma"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pVT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TCTTCCCAATGTGAAGTGGACCGCCGATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."  
 Soares and M.Fatima Bonaldo."

## FEATURES

source

## BASE COUNT

ORIGIN

Query Match 9.9%; Score 233.2; DB 9; Length 564;  
 Best Local Similarity 65.1%; Pred. No. 1.4e-51;  
 Matches 356; Conservative 0; Mismatches 188; Indels 3; Gaps 1;  
 QY 371 agcatggcgtgacccgcgattgcaggacatgttcgcgcgtattaatgaactcaga 430  
 DB 545 ACCAGGGCATATCCACAGCTCTGGAGAGACTCTTCTCGGATCAACAGCAGACCA 486  
 QY 431 aggaagaagaacctcaactgcacgcgtgaagtttcgtacttggaatttacaatgaacagag 490  
 DB 485 CG---ACANACATGTCTACTCCGTGAGGTGACGTACATGAGATTACTGTGAGCGCG 429  
 QY 491 tcggagactgtgaatcgtcgacaaaaggggaatctcaaggtccgagcaacacccgtcga 550  
 DB 428 TCGGTGACCTCTNTGACNCCAAAGAACAGGGCAACCTTCGCGTGGAGGAGACCCACTGC 369  
 QY 551 ccggccctcactgaggaggttggcgaagctgtgctgcgtatccattccaagaatacgaaa 610  
 DB 368 TGGGGCCTTACGTGGAGGAGCTCTCCAACTGGCTGTACCTCCTNATCATGATCATCCAGG 309

QY 611 atctcatgatgagggaacaaagccagaacggttgcgcgcacacaaacatgaacgagacat 670  
 DB 308 ACCTCATGACTCAGGGAACAAGCCAGGACCGTGGCGGCCACCAACATGAATGAGACCA 249  
 QY 671 ccagtcgatcccaacgcgctcttcactttgaccttgacgcgcaaaagtgcgatgaagaga 730  
 DB 248 GCAGTCGCTCCACGCCGCTTCAACATCATCTTACCCAGAAAGCCCATGACGCAGAGA 189  
 QY 731 ccaaaatggacacagagaaggttcgaagatcagctctggtagatttggcgggttctgagc 790  
 DB 188 CCAATATCACCCAGGAGAGGTGAGCAAAATCAGCTGTGGACCTGGCTGGGAGCGGAGC 129  
 QY 791 gagcaacgtccacccgagctacttgagcgagctgaaggggtgcagagatcaaacgcgt 850  
 DB 128 GGCGTGACTCCACGGGAGCCAGGCGCAGCGCTCAAGGAGGGGGCCCAACATCAACAAGT 69  
 QY 851 cacttcgacacctagctcgtgtgattgcagcgctagcgatagtcgtcgtcgggaaaacaga 910  
 DB 68 CGCTGACCACTGGGCAAGGTCTCTCCGCCCTGGCTGAAATGGACTCCGGACCCCAACA 9  
 QY 911 agaagaa 917  
 DB 8 AGAACAA 2  
 RESULT 14  
 BE252943  
 LOCUS  
 DEFINITION  
 601117534F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3358016 5', mRNA  
 BE252943  
 mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 BE252943.1 GI:9123079  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 586)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM166 row: c column: 09  
 High quality sequence stop: 586.  
 Location/Qualifiers  
 1..586  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3358016"  
 /clone\_lib="NIH\_MGC\_16"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."  
 BASE COUNT 136 a 162 c 168 g 120 t  
 ORIGIN  
 Query Match 9.6%; Score 226.4; DB 10; Length 586;



Job time: 3142 sec

---

**THIS PAGE BLANK (USPTO)**